INSTITUTE OF MATHEMATICAL GEOGRAPHY

MONOGRAPH SERIES

Nystuen/Dacey Nodal Analysis

Edited by

Keith J. Tinkler

Monograph #7

IMAGE IMAGINATION

MICHIGAN DOCUMENT SERVICES, INC.

NYSTUEN/DACEY NODAL ANALYSIS

Edited by

Keith J. Tinkler
Department of Geography
Brock University
St. Catharine's, Ontario
Canada.

with contributions by

Keith J. Tinkler
John D. Nystuen
Michael F. Dacey

Institute of Mathematical Geography

Michigan Document Services
Ann Arbor, Michigan
1988

Acknowledgments

My primary thanks are to Dr Sandra Arlinghaus of IMaGe whose constant encouragement has enabled this monograph to see the light of day. Her comments, and those of her referees, have led to a considerable improvement in the preparation of this monograph. Loris Gasporotto drew Figures 1 and 3, and Brock University has provided the computer support over many years which has enabled the number crunching and the word processing that are evident in this work. Peter Van Asten of the Computing Centre quickly and efficiently resuscitated the FORTRAN program (Appendix B).

ISBN: 1-877751-14-6

CONTENTS

"FOREWARD"

Graph theory is a mathematics of structure. Many years ago we (Nystuen and Dacey, 1961) used some theorems from graph theory in the analysis of the spatial structure of a nodal region. The application was novel. More commonly, Euclidean geometry had served as the means to formalize spatial aspects of geographical problems. Tinkler's work is an exception. Over the years he has developed graph theory methods of geographical analysis (involving nodal regions) that have gone well beyond our modest beginning. The present work shows the evolution of nodal analysis. Tinkler's paper of 1979 is of particular interest; in it, he draws from graph theory and probability theory to characterize message flows in a network of cities. He develops notions of the expected nodal structure and expected nodal flows. He identifies several types of nodes which may exist and provides several statistics suitable for nodal analysis. The paper presents logical development of several stimulating ideas and suggests others for further work.

Recent developments in computer-aided geographical information systems have created great interest and activity in geography and allied disciplines. I have noticed that the huge capacity for data management made possible by these systems has not been matched by particularly sophisticated structural analysis. Therefore, Tinkler's volume on fundamental spatial structure is timely and welcome.

John D. Nystuen

Professor of Geography and Urban Planning

University of Michigan

Ann Arbor, MI 48109

June 24, 1988

Preface

It is usually expected of an Editor that he explain the reasons for the work with which he burdens the world. I shall be no exception, but as I am something of a historian manqué in geomorphological matters I feel that an historical view of the origins of this monograph will eventually be of interest as an explanation, rather than a justification, of its blatant palaeo- and neo-structuralism - I suppose that is the proper epithet to apply to such work as this mongraph contains.

The story undoubtedly begins with the publication of the Nystuen-Dacey paper (here reprinted) that founded formal nodal analysis (though not nodal regions as an idea), and was also a 'graph theory in geography' first. As far as I can tell this has no real roots in other applied work, although I am unsure of the degree to which it found its spiritual roots in graph theory itself: certainly digraphs (which a nodal graph certainly is) were well described by 1961 in the mathematical literature. What is unusual is the method by which the digraph is abstracted from a complex data table. Since then the technique has been adapted in many ways, most of which are briefly mentioned in my 1979 review, the relevant part of which is reprinted here in Appendix A (item 1).

A characteristic of geography in the 1960s was its wholesale adoption of existing statistical techniques; usually from other social sciences and often pyschology. Even graph theory itself was no exception, although the method was not inherently statistical. But Nodal Analysis *was* different because it structured geographical data with a mathematical technique and in doing so provided a structural analysis. In terms of the complexity in the original data - an n x n matrix where n was usually well over 10 - an enormous statistical simplification was achieved, indeed by the order of $(1-n^{-1})$. Now it is tempting to think that in so doing one is throwing out the baby with the bathwater, that such an enormous reduction in complexity must inevitably lose much of significance.

Although philosophers praise simplicity and the beauty of Occam's razor, seasoned practitioners (and certainly sedulous critics out for an easy target) are all too ready to assume that it *can't* be *that* simple. Perhaps they are right, perhaps not, and it is not my purpose to debate the issue at length. However, one can point out that one reason that many statistical operations lost favour was because of the difficulties of providing good, and binding interpretations. Factor analysis is perhaps the prime example, but papers using graph theory also fell readily to the temptation to repeat what the last analyst had said, without any real attempt to link the processes being performed on the graph to the geographical world being modelled. Thus the interpretation connecting the theory to reality was missing. But something else was missing too. In relation to real world data sets (transaction tables in the present case) the discipline as a whole failed dismally to even see the point at which statistical questions about theoretical analysis really should enter: how does the statistical structure of the transaction table affect the answers that various techniques applied to the Table yield? The question is just as valid when asked about the data matrices being used for Factor Analysis (which I'm only picking on because it is probably the best documented example in the literature.)

The reader will find that in my review (Appendix A, item 2) of two papers that appeared after my 1979 review (Appendix A, item 1) there are echoes and reflections of these disciplinary flaws and failures. It seems to have been all too easy to drop a method and move on to something else which, in the end, proves to be no more satisfactory.

I would have no basis for carping about the deficiencies of the discipline, or even a tiny quantitative subset of it, unless I could claim to have made some efforts to remedy it myself. Whether my efforts are worthwhile is besides the point: but I can and do claim to have tried to make some amends in the limited field of graph theoretical analysis, and this monograph is merely the latest of these attempts, and I am very grateful to IMaGe for the opportunity to at least put the statistical analysis in Nodal Analysis before the 'public.'

I was trained as a geomorphologist, but from my very first University appointment I was required to teach Quantitative Methods to all and sundry. When I arrived at Makerere University College (University of East Africa as it them was, at Kampala, Uganda) in June of 1968 I was followed very shortly afterwards by Don Funnell, a graduate of Cambridge, who came to Makerere with the intent of working for a Doctorate on some aspect of Uganda's spatial economy. In the absence of suitable alternatives I was asked to act as a quantitative supervisor, and Funnell's interests soon ensured that I come to terms with the Cambridge approach to human geography, especially as embodied in Haggett's *"Locational Analysis in Human Geography."* In my subsequent readings I was appalled by the lack of a sound, or well justified, interpretative basis for the procedures being followed in the graph theory applications that I encountered. Also lacking was a general theoretical framework into which the techniques being used might fit. The lack was in itself excusable up to a point, but by the late 1960s it did not seem that serious efforts were being made to remedy the situation (despite the appearance of *"Network Analysis in Geography"* for example); and I might add that this was true of virtually *all* the techniques being applied, not just graph theory.

It was from this background that my 1972 papers on the interpretation of eigenvalues and a theory of radial graphs emerged. The next year my paper connecting graph colouring and rural periodic markets appeared, together with a paper that addressed a theory of the region. In the following years my move to the United States and then two years later to Canada interrupted the flow of papers along these lines although two papers that looked at statistical aspects of graphs were given at Conferences: the 1976 one on the expected distribution of nodality (given in Ann Arbor at a regional AAG meeting and reproduced in this monograph), and one in 1978 (given *in absentia* at the AAG in New Orleans) which sought to characterize all graphs which have the same local structure (i.e. which have identical labelled adjacency matrices, but different patterns of connectivity).

During these years I was also considering the problem of Nodal Analysis: specifically how could you find out the expected number of nodal regions that a given transaction table should have? At least as a start, that was the most obvious question to ask.

If you don't have a good, or any (statistical) theory, then brute force computer simulation is one avenue by which to obtain some hints that may lead to a solution: at the very least it gets you started! So I wrote some programs that would do Nodal Analysis on random matrices and was rapidly appalled by the quantity of output and the problems of reducing it without graduate student labour. Still I felt that this was the only way to go and decided, against my inner instincts which told me it was a total waste of public money, to apply for a research grant (with student labour) to tackle the problem. I sat down to write the grant application and in the process developed the paper which forms the core of this mongraph: there was then no need for the grant! I don't recall many of the details now, but once I realised that an estimating matrix, [M] in the monograph, could be based on the column sums of [T], the rest was mainly a matter of following my nose and applying standard principles, although sometimes these were used in relatively ingenious ways: e.g. the use of Markov techniques to generate the expected structure of nodal paths longer than 1. It does come to mind that the real insight came right after what looks like just a nifty trick: the use of a permutation matrix to re-order the initial Transaction Table according to the rank on the column sums. After this step nodal flows are those that occur below the main diagonal, and it then becomes visually obvious (if you are in the correct mind-set) that an estimation matrix like [M] need only consist of elements in those positions.

I would re-iterate that the re-ordering by rank order using the permutation was never intended to be anything other than a 'tidying up' trick, but once done something previously implicit then became explicit. Because I already had most of the 'parts' I needed for the FORTRAN program (markov chains, short path algorithms, powers of matrices, standard Nodal Analysis) that aspect of the work fell into place rapidly. The paper was then given at the *Canadian Association of Geographers* meeting in Montreal (May 1979), and the following January at the *Institute of British Geographers* meeting in Leicester, U.K. (in both cases to minuscule audiences!).

Since then it has languished, partly because I had other projects, and partly because it seemed unlikely that it would attract much use or attention. Perhaps, too, a lack of graduate students who might have been bullied into using it, and of colleagues of like mind, helped to keep it suppressed! It is only appearing now because I saw the IMaGe advertisement in the AAG newsletter in 1986. Because I connected John Nystuen with Michigan, which is the base for IMaGe, I wrote a letter of enquiry suggesting that it might be a very appropriate place for the paper to appear. After more delays which are almost entirely of my making, here it is!

The FORTRAN listing which is provided in Appendix B was retyped from a Xerox of the old program listing. It was uploaded to the Burroughs at Brock University, and after the typos were debugged it compiled and ran on a FORTRAN 66 compiler (old FORTRAN IV, which is what I originally used). The output agreed, decimal by decimal, with that on my 1979 output, still extant. The complete output from the input file for Kenya is given, except that the spacing has been adjusted for printing purposes. The program as listed can be provided by IMaGe as a Textfile on a Macintosh, IBM PC, or APPLE II (DOS 3.3 or ProDOS) disk.

In its simplest and most natural form the program performs a standard Nodal Analysis (including the re-ordering step to 'tidy up' the matrix), and computes a short path matrix for the nodal structure. Simultaneously it takes the column sums of the original transaction matrix and uses these as a basis for the statistical estimation of all the items needed in the analysis. Thus, at this step, the analysis is entirely internal. Two alternatives are provided: (i) to use a rank-size model (with exponent -1, though this could be changed easily enough) as a basis for the column estimates, or (ii) to use instead of the vector of columns sums, a vector of weights supplied by the user, and which presumably reflects some other measure of importance over the nodes of the system: perhaps population figures. A further alternative would be to supply an entire matrix of estimates: one for each ij element in the transaction matrix, and these might, perhaps, be derived from a variety of gravity models. The program would be easily modified to do this (indeed such a version was once tried and tested) but at the time it seemed too like special pleading: it provides an estimate for each potential nodal flow, and so it was abandonned in the final version.

In most applications one will be begin with the transaction table, induce the nodal structure, and then assess it via the statistics produced by the program. Still, it is worth re-iterating that the whole probability structure of a nodal analysis is implied by just *one* item defined over a set of nodes: a set of weights, which of course contains within itself an implicit ranking by virtue of its numerical attributes. Thus one can speak of the probabilitisic nodal struture of a set of cities. Because this makes no direct reference to the spatial arrangement of these nodes it contains spatial attributes only to the extent that the set of weights does. *If* the spatial economy produces *puckers* in the nodal system, as reflected in the actual nodal structure revealed by the transaction table in use, *then to this extent* the present system of analysis reveals the impact of additional localised geographical factors, over and above those which operate to produce the weights. The text describes an example of this in the way that Eldoret *draws* an actual nodal flow from Kitale which on probabilistic grounds is extremely unlikely.

Keith Tinkler

Keith Tinkler

References

Nystuen, J.D. & Dacey, M.F., 1961. A graph theory interpretation of nodal regions. *Papers and Proceedings of the Regional Science Association*, 7, 29–42.

Tinkler, K.J., 1972a. Physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, 55, 17–46.

Tinkler, K.J., 1972b. Bounded planar networks: a theory of radial structures. *Geographical Analysis*, 4, 5–33.

Tinkler, K.J., 1973. The topology of rural periodic market systems. *Geografiska Annaler*, 55, 121–33

Tinkler, K.J., 1973. A theory of the region. *East African Geographical Review*, 11, 5–22.

Tinkler, K.J., 1976. *The expected distribution of nodality in random (p,q) graphs and multigraphs.* Paper presented to the East Lakes Division, Association of American Geographers, 16 October 1976, Ann Arbor. [reprinted in this monograph without being refereed]

Tinkler, K.J., 1978. *Characterizing graphs of networks which all have the same local structure.* Paper presented to the Association of American Geographers, meeting in New Orleans, April 1978.

Statistics for Nystuen–Dacey Nodal Analysis

Introduction

When Nystuen and Dacey (1961, hereafter ND61 for brevity) introduced the method of nodal analysis into Geography for the purpose of delineating a certain category of functional region on the basis of a table of transactions between a set of places, they devised a method of considerable elegance and originality for a specifically geographical problem. The technique has attracted considerable use over the years (for a full review see Tinkler 1979, the relevant section is reprinted in this monograph together with references, Appendix A), and while it may be criticised for using only an apparently small part of the total information in a typical transaction table, its virtue is exactly that – it achieves considerable simplification in what is often a complex system, greatly aiding interpretation.

Somewhat surprisingly perhaps, while the method has been readily adopted as an empirical technique to structure a transaction table, few people if any, have addressed the abstract question: how expectable are the results? For example, how many nodal regions should there be in an $n \times n$ table? To what extent would such a result depend on the frequency distribution of the elements in the table? How many nodes should each nodal region contain, and what is the expected maximum distance a node in a region would lie from the focal or terminal city? To my knowledge no attempts have been made to consider these questions, and the aim of this paper is to develop and illustrate a statistical methodology that is capable of answering them. Given a square transaction table representing the interactions amongst a set of places and with self-interactions at each place omitted from consideration (the main diagonal of the table), the procedure developed below derives all the required statistics from the entries in the transaction table, and more particularly from their column sums.

However, just as the nodal scheme itself may be worked on the basis of some external ranking of the places in the transaction table (ND61), so the required statistics also may be based on an external set of values (weights, in addition to ranks) for the places. In this way it becomes possible to test an empirical nodal system against the expected nodal structure of a chosen geographical model, with a consequent improvement in the sharpness of interpretations. For example, it would be possible to use a set of weights for places based on some variant of the rank-size rule – a hoary old geographical favourite.

In a more specific direction the methodology can be adapted to deal with the situation where the elements in the transaction table are regarded simply as estimates of the true values. When they are suitably normalised, row-wise as in a transition probability matrix they can be used to estimate the probability that a certain transaction occurs. When they are normalised over the entire matrix they can be used to estimate the probability that a certain nodal structure occurs.

Significance testing

The issue of testing the statistical *significance* of the expected values obtained from the methods in this paper, vis-à-vis observed values obtained by standard Nodal Analysis of an empirical transaction table, is not touched upon. In most cases an application of a Chi-Squared test would be appropriate, but for small tables there may be difficulties with small expected frequencies for which corrections may have to be made.

Restructuring Nodal Analysis

The mechanics involved in determining the matrix of nodal flows, [N], from the transaction table or matrix [T], are well known (ND61) but the method will be reviewed briefly. It is immaterial to the discussion in this paper whether the matrix [T] has been adjusted beforehand, via powering procedures, to take account of indirect flows, or not. Such procedures are detailed in ND61 and they are discussed at length in Tinkler (1976). It is possible though that adjustments of that type might complicate the process of interpretation unduly.

The column sums of [T] are first summed to establish a ranking of places according to the number of incoming transactions they receive. (Nystuen and Dacey remark that an external set of weights can be used to rank the places according to criteria such as population size or central function. However, no examples of the use of this alternative method are known to me.) In each row the largest element is marked. This element, indexed by the i^{th} row and the j^{th} column (and excluding the case $i=j$), is marked as *nodal* and is entered with a 1 in [N] *if and only if* the i^{th} column sum is smaller than the j^{th} column sum, and with 0 otherwise. Thus a flow is *nodal* if the place to which it is going is ranked higher in the hierarchy than the place from which it is coming: places of high rank therefore receive more transactions, and nodal flows, than do places of low rank.

For the purpose of analysis it is advantageous to restructure this procedure by reordering [T] according to the rank of its nodes on the columns sums (or according to an external criterion if that variant of the method is used). For example, given:

$$[T] \quad = \quad \begin{bmatrix} - & 0 & 6 & 3 \\ 3 & - & 4 & 3 \\ 5 & 6 & - & 3 \\ 4 & 2 & 3 & - \end{bmatrix}$$ (largest element in each row is boldface)

Column Sums	12	8	13	9
Rank	2	4	1	3

it can be reorganised to give:

$$[T]^{*} \quad = \quad \begin{bmatrix} - & 5 & 3 & 6 \\ 6 & - & 3 & 0 \\ 3 & 4 & - & 2 \\ 4 & 3 & 3 & - \end{bmatrix}$$

Column Sums	13	12	9	8
Ranks	1	2	3	4

In this procedure, and throughout this paper, a numerically low value for the rank signifies a high ranking node; in this way the row and column indices of $[T]^{*}$ indicate the rank of the nodes on the column sums of $[T]$.

It is assumed that no ties occur in the ranks based on the column sums, or among the row elements insofar as they affect the determination of the maximum element. Ties within rows can usually be resolved by using the maximum flow corresponding to the element with the highest column total. If ties occur in the column totals, then insofar as they affect the assignation of nodal flows then the tie must be resolved on the basis of external criteria, although an alternative might be to consult the outgoing flows for the nodes concerned: the one with the smallest outgoing flow might be thought to be more of a sink in the regional system.

The reordering of [T] to give [T]* can be accomplished by calculating:

$$[T]^* = [P]^T[T][P],$$

where [P] is a permutation matrix and $[P]^T$ is its transpose. [P] can be established from the vector of ranks for the column sums of [T]. If the j^{th} column [T] has rank k then:

$p_{ji} = 1$ for $k = i$, $= 0$ otherwise.

In this example, therefore, [P] is found to be:

$$[P] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

It is now possible to reduce [T]* to a matrix [F] in which the largest element on each row is marked 1 in [F] where an element in [T]* is boldface, and is marked 0 otherwise. The result is:

$$[F] = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Now, because of the strict requirement that nodal flows may only proceed up the hierarchy, (to nodes with larger column sums), it is possible to show a *potential* nodal flow matrix [PN] as follows, in which a 1 shows that the ij^{th} entry is *potentially* nodal, and a 0 that it isn't.

$$[PN] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Because of the restructuring, [PN] has the form of a lower triangular matrix. The matrix of nodal flows [N] follows directly as :

$$[N] = [F] \square [PN] \qquad \text{........(1)}$$

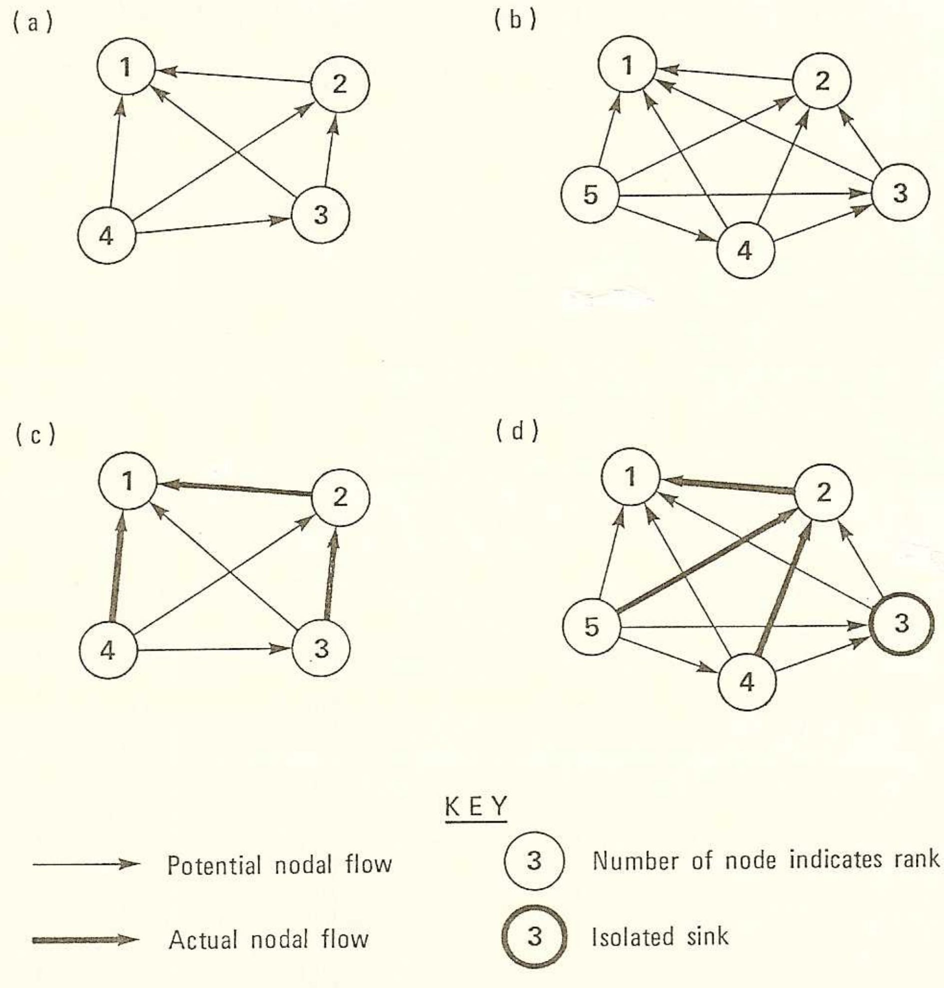
where □ indicates element by element multiplication of the [F] and [PN] matrices; i.e., the same formalism as is used in ordinary matrix addition and subtraction). Thus:

$$[N] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \square \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad ...(2)$$

and [PN] acts as a filter which only permits flows up the hierarchy. It is the lower triangular part of [F] that yields [N]. As a graph, [PN] is represented in Figure 1 (a,b) which makes it clear that long indirect paths through the nodal hierarchy are only possible for nodes which rank relatively low. It is the existence of this strict ordering which makes possible the probability analysis given below so readily computable (with the aid of the appended FORTRAN support; Appendix B). In addition all possible nodal

FIGURE 1

Potential and actual nodal structure in a small example (see text).



structures in [N] are sub-graphs of [PN] with the condition that a row sum of [N] is either 0, or 1. Figure 1 (c) shows the graph [N] embedded in [PN] for this example. Thus, a unique nodal flow exists out of the i^{th} node, or it does not. There is no such restriction on column sums. Nystuen and Dacey (1961) have already shown that (1) their procedure partitions the set of places in the system and (2) that each nodal component has a unique terminal point which I will here term a *sink*. These results are taken as understood and it follows that counting sinks will serve to count the number of nodal regions in the system. This in turn is equivalent to counting the rows in [N] which have a sum of zero; i.e., have no nodal outflow.

In addition to counting the expected number of nodal regions (sinks), other statistics that will be of interest include: (1) the expected number of nodes tributary to the i^{th} node at k steps, as an estimate of the expected size of each nodal region. (2) The probability that any particular node will serve as the sink of a nodal region and (3) the expected numbers of different node types. These types are: (i) sinks which have tributary nodes (ii) isolated nodes with no tributary nodes (trivial according to the terminology of ND61) (iii) source nodes, which have no incoming nodal flows and, (iv) passing nodes which both receive (possibly several) nodal flows and which generate one outgoing nodal flow (by definition, see above).

Because the highest ranking node is a guaranteed sink, particular interest may focus on the nodal region tributary to this node.

Approaches to the statistical analysis of nodal structure

From equation (1) the only apparently variable element in the analysis is the matrix [F] showing the position of the maximum flow in each row. However, there is unfortunately a hidden subtle effect in [F] because even if the elements of [T] are chosen from (say) a uniform distirubution of random numbers it is nevertheless true that the column sum which is largest in the resulting [T] matrix has a greater chance of having more of the maximum (and for this particular column, nodal) flows than has the column with the smallest column sum. When [T] is reorderd according to the size of the column sums to give [T]* then from equation (2) it may be seen that this causes the nodes with high column sums to be much more likely to receive nodal flows, since there are many more 'smaller' nodes (lower down the hierarchy) which potentially can send nodal flows to it. This bias may be restated by saying that in the matrix [F], which is based on [T]* (the reordered matrix), more of the maximum flows occur below, than above, the main diagonal.

Nodal Analysis in the limiting case of random transaction tables

In the case of random transaction tables this bias can be ignored, or rather its effect can be reduced, by considering the limiting case as n, the dimension of the (square) table increases without bound. It is then possible to argue that the probability that the j^{th} column in the i^{th} row is a nodal flow is $(n-1)^{-1}$, (remembering to exclude the diagonal position). The Central Limit Theorem can then be applied to the column sums and as the limit is approached the variance amongst the sums will tend to vanish and although a ranking may be established amongst the column sums on the basis of tiny differences in magnitude, the nodal flows will be virtually independent of the ranking on the column sums. For small n this situation can be simulated by building a random transaction table where the ij^{th} element is constructed by $(1+í)$ where í is a random variable and $í \ll 1$. Either of these situations is clearly very unrealistic but it is worth pursuing the argument because with only minor modifications the more realistic analysis follows identical methodology. Let $m = (n-1)^{-1}$ be the probability that the ij^{th} element of [T]* is a nodal flow, then:

$$[M] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ m & 0 & 0 & 0 \\ m & m & 0 & 0 \\ m & m & m & 0 \end{bmatrix}$$

This matrix, [M], is the probabilistic analogue of [PN]. The aim is to establish the expected number of sinks, or rows with sum = 0. The highest ranking place (the one with the largest column sum) can have no outgoing nodal flow since there is higher ranking place to which it might be sent. In [M] this is represented by the first row which has sum zero, with probability 1. Because we know that each row of [M], in a realisation of the process, can have only *one* nodal flow then we can deduce from [M] by standard probabilistic arguments that the probability that the i^{th} row has a 1 is $m(i-1)$, and that it has sum zero is $1-(m(i-1))$. The total expected values can be found by summing either of these terms over all the elements. It is perhaps easiest to count the number of nodal flows directly and then subtract these from n to obtain the number of sinks. In a lower triangular matrix of dimensions *n x n* there are $(0.5(n^2-n))$ entries and remembering that each element is equal to $(n-1)^{-1}$ then the expected number of nodal flows is:

$$\text{expected number of nodal flows} = \frac{n^2-n}{2(n-1)} = \frac{n}{2}$$

which will also be equal to the number of sinks, because these are the complement to nodal flows. The variance on this mean may be calculated by the usual methods but as the result is only of limited interest I shall simply report the fact that the exact value for the variance is, where $k=n-1$:

$$\text{var(expected number of sinks)} = \frac{14k}{3} + \frac{7}{3k} + \frac{2}{k^2} + 1$$

For large n this is well approximated by the first term alone, so that for a 1000 node random transaction table there would be about 500 expected sinks with a standard deviation of about 68.

Returning to [M] the expected number of nodal flows incoming to the j^{th} node can be computed by adding up the entries in the j^{th} column. In the limiting case then the j^{th} ranking node will have $[(j-1)m]$ expected incoming nodal flows, a probability that must decline for the lower ranking nodes until the lowest ranked nodes is reached, a node which cannot receive a nodal flow.

Because of its general inapplicability it is not worthwhile pursuing this limiting case any further in detail (we don't deal with large random transaction tables, or even approximations to them!). The logic reveals that the probabilistic matrix [M] is all that is required to generate the required statistics and the only problem is to find a more realistic distribution for the entries with respect to a given empirical problem.

Nodal Analysis of empirical transaction tables: direct flows

An element in [M] estimates the probability that a particular nodal flow is present. In realistic systems we would expect field effects, generated by the larger nodes in the system, to attract flows to those nodes preferentially. Table 1 shows the pattern of telephone calling amongst ten Kenyan cities during July 1967.

TABLE 1

Kenya: telephone trunk (long distance) census, July 1967 (5 days 30 hrs)

from / to	1	2	3	4	5	6	7	8	9	10
Eldoret 1	0	14	79	189	1	172	10	166	0	6
Kericho 2	33	0	222	8	8	396	14	270	1	0
Kisumu 3	121	148	0	32	35	757	3	188	2	5
Kitale 4	265	3	56	0	7	233	0	88	0	5
Mombasa 5	0	3	0	2	0	1315	1	33	0	2
Nairobi 6	284	230	633	110	1747	0	386	1153	334	577
Naivasha 7	6	7	1	2	2	420	0	272	0	1
Nakuru 8	109	190	191	69	30	1800	209	0	21	16
Nanyuki 9	4	0	2	1	2	410	1	76	0	221
Nyeri 10	4	1	2	1	10	805	1	37	183	0
Σ	826	596	1186	414	1842	6308	625	2283	541	833

Total sum over all the table, 15454 calls

The powerful effect of Nairobi (the capital) is obvious, while smaller field effects are displayed by Mombasa and Nakuru, and to a small extent by Kisumu. One reasonable way to allow for this in our revised version of [M] is to assume that the ijth element in [M], $(i > j)$, is estimated by the jth column sum (which represents its drawing power in transactions) over the sum of all column sums (all transactions in the system). For Nairobi this yields (6308/15454) = 0.4082. However, this value cannot be applied directly to the same column in each row because in each row one diagonal element is missing. In consequence each jth row of [M] has to be adjusted by multiplying it by $(1-p_j)^{-1}$, where p_j is the jth element in the probability vector derived from the column sums in Table 1 according to the prescription just given. Table 2 shows the [M] matrix obtained in this way with the nodes reordered (as described earlier) on the basis of the values obtained from the column sums in Table 1. The probability vector derived from the column sums in Table 1, and used to construct it is also given in Table 2. Note that in each column the values decline downwards. This is because the value of $(1-p_j)^{-1}$ is declining to 0 as smaller and smaller elements are encountered for p_j, the pattern is readily visible thanks to the re-ordering process.

The method might be regarded as somewhat deficient because it assumes that all nodes are affected by each other's fields in essentially equal strengths, after allowances of the type detailed in the last paragraph are taken into account. On the other hand a more detailed specification of [M] would come close to simply reproducing [T]*, which would serve little useful purpose. The implications of such a procedure will be considered later but the method advocated, based on the column sums, appears to be a reasonable one in that it takes into account all flows in the system, and the way that they are distributed amongst the nodes. Assuming the [M] matrix is reasonable I shall now show how it may be used to compute the required statistics using the Kenyan example.

TABLE 2

The matrix [M] based on the transactions from Table 1

Note that the matrix has been reordered by ranks

rank		1	2	3	4	5	6	7	8	9	10	
	old index	6	8	5	3	10	1	7	2	9	4	place
1	6	----										Nairobi
2	8	0.479	----									Nakuru
3	5	0.463	0.168	----								Mombasa
4	3	0.422	0.160	0.129	----							Kisumu
5	10	0.431	0.156	0.126	0.081	----						Nyeri
6	1	0.431	0.156	0.126	0.081	0.081	----					Eldoret
7	7	0.425	0.154	0.124	0.080	0.056	0.056	----				Naivasha
8	2	0.425	0.154	0.124	0.080	0.056	0.056	0.042	----			Kericho
9	9	0.423	0.153	0.124	0.080	0.056	0.055	0.042	0.040	----		Nanyuki
10	4	0.419	0.152	0.122	0.079	0.055	0.055	0.042	0.040	0.036	----	Kitale

The matrix is based on the probability vector:

0.408 0.148 0.119 0.077 0.054 0.053 0.040 0.039 0.035 0.027

which is based on the vector of column sums of Table 1.

Expected number of sinks

The method is identical to that discussed for the random transaction tables. The probability that the i^{th} row has a nodal flow is found by summing the values in the i^{th} row, $\Sigma i\blacksquare$, where $\blacksquare$ indicates summation over the full range of the index in that position. Summing all the row totals obtained gives the expected number of nodal flows, which is equivalent to summing all the elements in the matrix below the main diagonal:

$$\text{exp(sinks)} = \sum_{i=2}^{i=n} \sum_{j=1}^{j=i-1} m_{ij} \quad \text{or } \Sigma m\blacksquare\blacksquare \qquad ...(4)$$

Subtracting this value from n gives the expected number of sinks or nodal regions. Table 3 shows the results for the [M] in Table 2. The total number of expected nodal flows is 7.291, so that the expected number of sinks is 2.709. No attempt has been made to compute the variance, although this could be done by standard methods, since the frequency distribution of sinks is likely to be strongly skewed so close to the origin, and indeed James Furst in a student project under my direction confirmed this for a sample of small random matrices.

TABLE 3

Sinks and nodal flows in Table 1

sinks		nodal flows	
Expected	observed	Expected	observed
2.709	2	7.291	8

One might use a Chi-squared test to test for differences between observed and expected frequencies, although there may be problems with small expected frequencies. In this case the two measures are close in any case: there are two actual nodal regions revealed by [T]* in Table 4 (the Kenyan towns), and 2.7 are expected (Table 2 & 3). When the observed is smaller than the expected then, whatever significance might be attached to the actual numbers involved, the system is tending in the direction of greater structuring - fewer nodal regions (see Figure 2, page 14, for a map of the nodal structure).

TABLE 4

[N], Nodal flows for the re-ordered matrix [T]*

Town		6	8	5	3	10	1	7	2	9	4	row Σ	col Σ
Nairobi	6	–										0	7
Nakuru	8	1	–									1	0
Mombasa	5	1	0	–								1	0
Kisumu	3	1	0	0	–							1	0
Nyeri	10	1	0	0	0	–						1	0
Eldoret	1	0	0	0	0	0	–					0	1
Naivasha	7	1	0	0	0	0	0	–				1	0
Kericho	2	1	0	0	0	0	0	0	–			1	0
Nanyuki	9	1	0	0	0	0	0	0	0	–		1	0
Kitale	4	0	0	0	0	0	1	0	0	0	–	1	0
Σ	7	0	0	0	0	1	0	0	0	0			

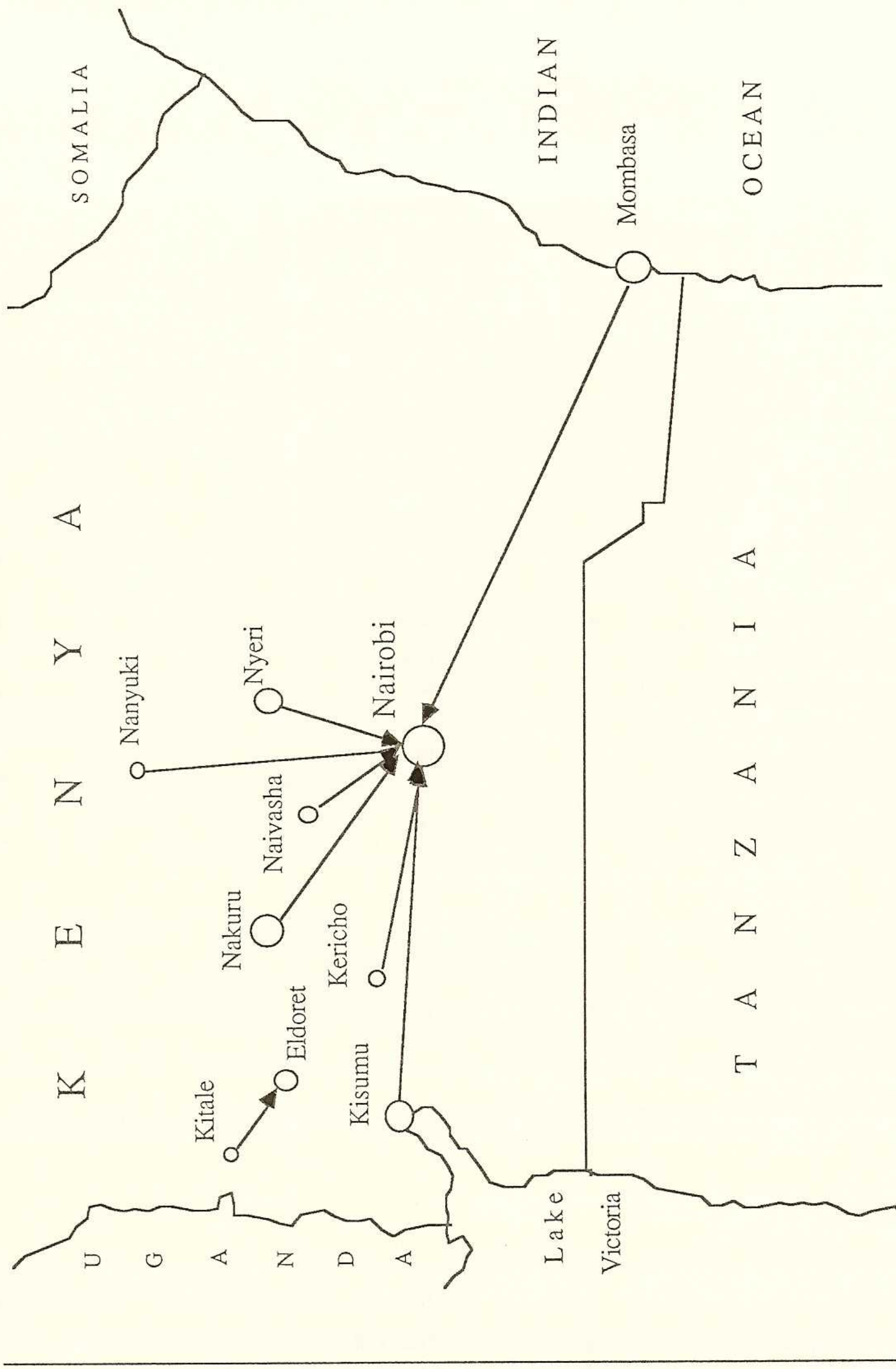


Figure 2 The nodal structure of Kenyan cities according to the telephone call census of July 1967, see Table 1, page 7 in the text.

Note: Circles for cities are roughly proportional to the column sizes in Table 1

Expected number of different node types

A more refined distinction between different node types was suggested in an earlier section. Possible row sums in the nodal matrix [N] are 0 and 1, and possible column sums are 0 or (positive) integer. Taking all four possible combinations for the i^{th} node yields the following structural types:

row (out)	column (in)	type
0	0	isolated sink
0	integer	sinks with tributary nodes
1	0	source nodes - no incoming flows
1	integer	passing nodes with incoming nodal flows, one nodal flow out

To count these expected frequencies for the i^{th} node requires the evaluation of the probability of the appropriate row or column sums. To count isolated sinks we sum for all i nodes the product of the probability that the i^{th} row is zero with the probability that the i^{th} column is zero. This can be written as:

$$\exp(0\text{-}0)\text{ nodes} = \sum_{i=1}^{i=n} \left[\left[1 - \sum_{j=1}^{j=i-1} m_{ij} \right] \left[\prod_{j=i+1}^{j=n} (1 - m_{ji}) \right] \right] \quad .....(5)$$

Because isolated sinks are a subset of the set of all sinks, calculated in the previous section, then the category (0-integer) can be obtained by subtraction from equation (4) although of course a direct expression would be possible.

A similar strategy can be applied to the category (1-0); the relevant equation is:

$$\exp(1\text{-}0)\text{ nodes} = \sum_{i-1}^{i=n} \left[\left[\sum_{j=1}^{j=i-1} m_{ij} \right] \left[\prod_{j=i+1}^{j=n} (1 - m_{ji}) \right] \right] \quad .....(6)$$

Again, the (1-1) category is found by subtracting equation (6) from (n - equation 4).

TABLE 5

Nodal Structure based on four node types

type	observed	expected
0-0	0	0.897
0-integer	2	1.812
1-0	8	5.674
1-integer	0	1.616

Table 5 shows the observed and expected frequencies for the four categories. Perhaps the most interesting category here is the (1 - integer) or passing node because nodes of this type imply intermediate steps in the hierarchy of flows - a characteristic called step-wise migration in studies of population mobility. A deficiency for the observed frequency of this node type with respect to the expected would imply short-cuts in the hierarchical structure; the converse would imply step-wise processes at work in forestalling flows routed towards the eventual sink.

Expected number of inflows to the j^{th} ranked node

These quantities are very easily computed since they are the column sums of [M], and there is no restriction on the possible number of nodal flows coming into the j^{th} ranked node, other than that imposed by the ranking, which means that the j^{th} node has a maximum of (n-j) possible nodes tributary to it. Table 6 shows these values derived from [M]. In a regional system, interest will often focus on an obvious primary node and therefore a good test of structure might be to see whether the actual size of this regional structure (the number of tributary nodes) is larger or smaller than expected.

TABLE 6

Number of nodal flows into the j^{th} node
Kenyan Cities

Node	Observed	Expected
6	7	3.9395
8	0	1.2524
5	0	0.8752
3	0	0.4804
10	0	0.2804
1	1	0.2216
7	0	0.1255
2	0	0.0796
9	0	0.0360
(4	0	0 by definition)

The interest in Table 6 is twofold. Nairobi has about twice the expected number of tributary nodes and dominates the Table. On the other hand there is little expectation from the figures that Eldoret (node 1) will be a sink. On the basis of the recorded transactions in the Table it is Nakuru (node 8) that 'ought' to be the next largest sink, whereas it is the lowly Eldoret that actually achieves this status. Here is a clear indication of a geographical effect - Nairobi is drawing in all the nearby nodes, even Mombasa, whereas the isolated Eldoret, in the far north west, is able to act as a small local sink (this can be clearly seen on Figure 2, page 14).

Expectation that the i^{th} node has a nodal flow (i.e., an outflow)

These quantities are computed as the row sums of [M], and for the Kenyan example they are tabulated in Table 7. The first and last elements in the Table, 0 and 1 respectively, merely re-affirm the fact that one sink and one source are guaranteed by the structure of nodal analysis. The Table indicates that the probability that a node has an outgoing nodal flow is inversely proportional to its hierarchical rank in the system, though the exact values depend on the entries in [M] which are derived from the transaction structure. Once again Eldoret stands out as running counter to the trend; the probability that it has a nodal flow is 0.85, whereas in fact it does not have one.

TABLE 7

Probability that the i^{th} node has a nodal flow
Kenyan Cities

Node	Observed	Probability
6	0	0.0000
8	1	0.4789
5	1	0.6311
3	1	0.7312
10	1	0.7947
1	0	0.8512
7	1	0.8954
2	1	0.9357
9	1	0.9722
4	1	1.0000

Statistics for Nodal Analysis - indirect nodal flows

Actual short path structure

Because nodal flows are directed strictly up the hierarchy, any nodal region is a directed tree in the graph theory sense of the word (see Tinkler 1977 for an introduction to graph theoretical terms). Such a tree is very similar to a river system except that there is no topological limit to the number of incoming nodal flows to a node other than those set by the size of the table and any particular node's position in the hierarchy. The tree structure of the graph of the nodal flow matrix [N] ensures that within a nodal region there is exactly one directed path from any tributary node in the region to a sink. In consequence, a standard accessibility analysis of [N], obtained by powering [N] (see Tinkler 1977 for review) will be of little interest and it is better to go directly to the short path matrix [S], of [N]; [S] = s([N]). This may be done by powering, or more efficiently by using Floyd's algorithm, which is coded in the FORTRAN support (Floyd 1962). For the Kenyan example [S] is identical to [N] (Table 4) and so it will not be reproduced again, but in general this will not be so. The number of tributary nodes in any nodal region can be established directly by counting the number of non-zero entries in those columns of [S] which are sinks (i.e., which have row sums of 0 in [N]). The short path distances in the same columns can be used to construct the structure of each nodal region in terms of the number of nodes lying at successively greater distances from the sink. (In an $n \times n$ system with t sinks the maximum distance a node can be from a sink is (n−t), and most will much nearer to a sink than this).

Statistical short path structure

Statistical information on indirect routes, that is to say the probability of paths of length k in or out of a selected node, can be obtained by normalising [M] to the form of a transition matrix, (the notation •[M] will indicate that this has been done), and then manipulating it according to standard markovian methods (for the markov methods and terminology used here *see* Kemeny and Snell 1960). Three types of analysis are described below, and are available in the computer program where they are performed sequentially. Initially [M] must be normalised to have row sums equal to 1 for all rows with a sum larger than zero, this gives •[M]. The first two analyses can be performed on •[M]. For the third type of analysis •[M] must be *renormalised* after structural adjustments have been made which involve deleting columns, which in turn will affect row sums. This form will be marked by ••[M], and the notation •(•)[M] will indicate that a choice of either form is possible depending on what is required by way of eventual interpretation.

The reason for the initial normalisation is that [M] was established to estimate the first order properties of the system, and the upper triangular portion of the matrix was ignored because although it can be calculated in principle, it does not describe nodal flows and is therefore irrelevant. Analysis at step lengths greater than 1, however, pertains only to the nodal structure itself and therefore only permissible nodal flow paths are taken into account (those described by 1's in [PN]). Renormalising non-zero rows after the upper triangular part of the matrix is discarded (including the main diagonal) allows for this need. Rows with all elements equal to zero (the first row of an •[M] derived from a [T]* matrix always has this property) are equivalent to sinks or *absorbing states* in a markov chain. In fact in large part •[M] is the transient part (termed [Q] in Kemeny and Snell 1960) of the transition matrix of an absorbing markov chain, and this is true for all three types of analysis described below. The only differences from standard absorbing markov chain analysis are that (a) the focus of interest is on the column sums rather than the row sums and (b) the absorbing states, the sinks for each nodal region, are left in the matrix instead of being partitioned out. In neither case does this affect the analysis performed.

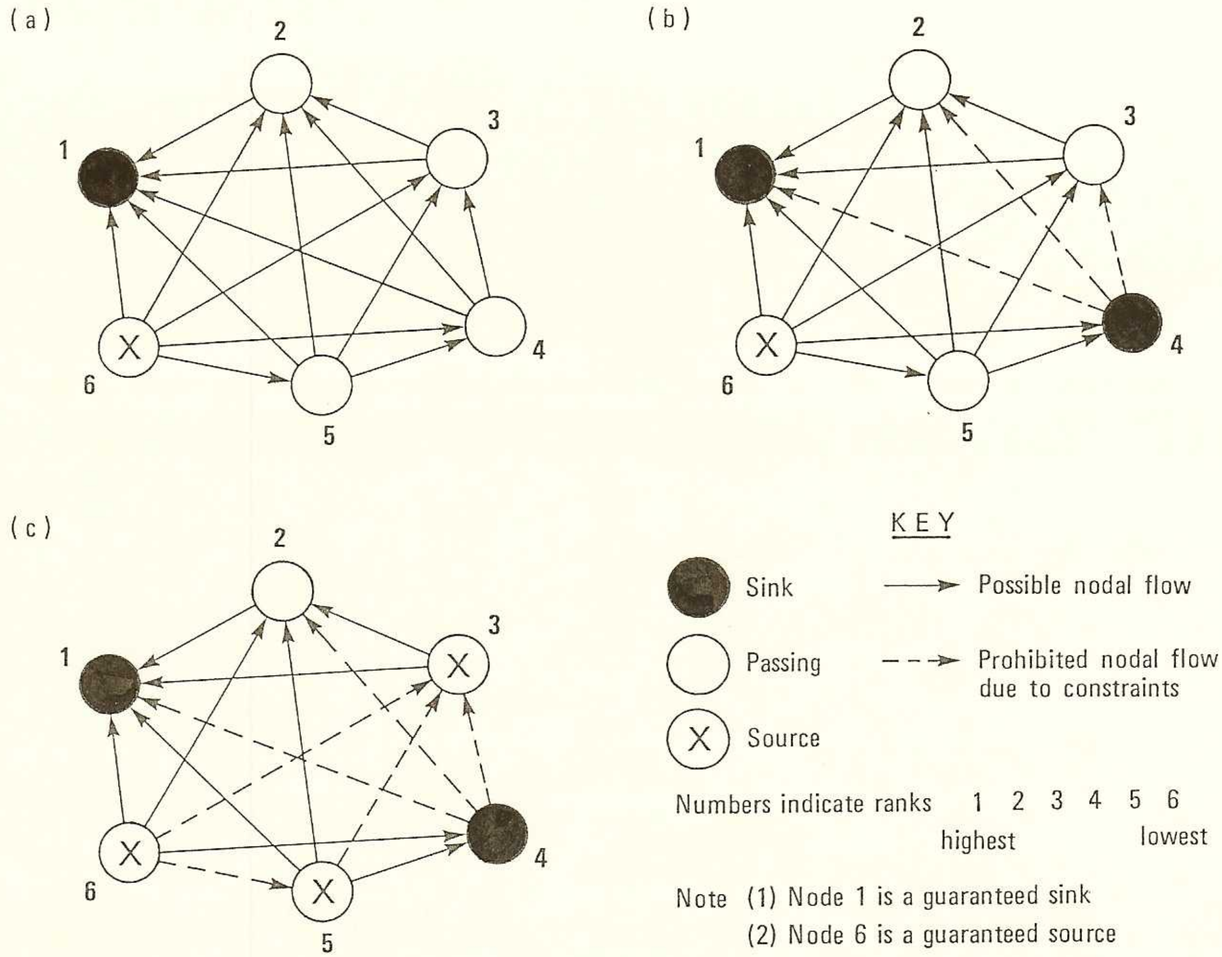
The expected number of k-step links for any k is found by computing the k^{th} power of •[M]. The interpretation of $\bullet[M]^k$ is the same as that for identical procedures applied to a normal accessibility or transition matrix (for details of these interpretations *see* Tinkler 1977 & Tinkler 1987). While full details may be obtained by inspecting the ij^{th} element of $\bullet[M]^k$, it will be sufficient for present purposes to consider the row and column sums of $\bullet[M]^k$.

Three types of indirect flow analysis can be made and I shall call them:

prime sink constrained (Figure 3(a))
actual sink constrained (Figure 3(b))
actual sink and source constrained (Figure 3(c))

FIGURE 3

Various constraints on indirect flow analysis in a small example.



The first type leaves *[M] as it is for the powering, and all flows eventually arrive (in a probabilistic sense) at the only sink - the first ranked node, which is automatically a sink, Figure 3(a). The second type specifies those nodes to be sinks which are found to be so from a Nodal Analysis of the actual system, i.e., the nodes with row-sums of zero in [N]. To cater to this the same rows in *[M] have all their elements set to zero. The effect of this is to interrupt some of the flows up the hierarchy so that no flows onward from known sinks are permitted, Figure 3(b). In this way the intervening effect of sinks on the path structure to the primary sink can be evaluated and compared with the results obtained from type 1 analysis.

The third type specifies not only the observed sinks but in addition specifies the observed sources - nodes which have zero columns in [N]. For these nodes the corresponding columns in *[M] must have their columns set to zero, and *[M] must be recalculated to give ●●[M]. In this case the only probabilistic paths available are either directly between specified sources and sinks or through such passing nodes as the system possesses, Figure 3(c).

Before proceeding to the uses to which the three types of analysis can be put it will be useful to review the results which can be obtained from $●(●)[M]^k$ irrespective of which type of analysis is performed, and this will aid in interpreting the output from the FORTRAN support.

Expected number of incoming k-step links to node j

This value is obtained as the column sum of the j^{th} column of $●(●)[M]^k$. The result may be compared with the observed number of k-step paths which is found from [S] by counting the number of entries equal to k in the j^{th} column. It is important to note that because of path uniqueness from any node to another in the hierarchical tree (this is true of any tree) counting the number of k-step paths to a node is the same as counting the number of nodes tributary to j at k-steps away, and which by definition lie further down the hierarchy. When k=1 the total number of expected paths must equal n minus the number of sinks specified, but the distribution of expectations by node may be revealing. For nodes which are sinks this provides a method of computing the observed and expected sizes of the nodal regions tributary to them. For other nodes it computes the number of nodes below them in the hierarchy, both expected and observed. The example, Table 8, shows the results for a k=2, Type 2, analysis for the Kenyan Cities.

TABLE 8

Type 2 Analysis
Incoming 2-step links to j^{th} node

Node	Expected	Observed
6	2.46	0
8	0.41	0
5	0.15	0
3	0.05	0
10	0.02	0
1	0.01	0
7	0.01	0
2	0.00	0
9	zero by definition	
4	zero by definition	

The final two entries are zero because a city must be three ranks up the hierarchy from the bottom to be able to receive an incoming two step link. The Table indicates that Nairobi ought to receive at least two incoming two step nodal flows, but fails to do so, an indication of the strong structuring effect that the city has on the system, even taking into account the existing transaction structure.

Expected probability of an outgoing k-step link from node i

This quantity, if it is required, can be found from the row sums of ●(●)$[M]^k$ and may be compared to the existence, or not, of such a k-step link in the i^{th} row of [S]. Because of path uniqueness in the nodal system there is only one node at k-steps up the hierarchy from i, or there is none. Table 9 is an example which complements Table 8.

TABLE 9

Type 2 Analysis
Probability of an outgoing 2-step link from the i^{th} node

Node	Expected	Observed
6	zero by definition (sink)	
8	zero by definition	
5	0.27	0
3	0.40	0
10	0.46	0
1	0.00	(Type 2 sets this as a sink)
7	0.46	0
2	0.49	0
9	0.51	0
4	0.53	0

It shows the probability that the i^{th} node has an outgoing two-step link (i.e., two successive nodal flows.) In an opposite fashion to Table 8 the first two nodes in the hierarchy cannot generate two step links up through the hierarchy, and by definition in Type 2 analysis the other sink, node 1, is defined as a sink, and therefore cannot generate nodal (out) flows. It can be seen that there is a roughly 50% chance that each of the nodes in the lower part of the hierarchy will generate a two-step link up the hierarchy, but they have failed to do so, being pulled directly into Nairobi with one step links. Thus the method shows exactly the manner in which the actual nodal structure departs from the theoretical expectation.

Total expected k-step links in the nodal system

Summing the results for all i or all j respectively in the last two sub-sections will yield this value which may be compared to the observed value in [S]. Divide this expected sum by n will give the probability that a k-step link exists in the system. When this value falls below some chosen value the powering of ●(●)[M] in a computer program may be stopped to save time (0.01 is used in the FORTRAN support). Otherwise ●(●)[M] will be powered to (n-1), although the probability of very long paths is extremely low and they are likewise very rare in actual nodal systems. As noted earlier their existence would imply the existence of step-wise processes.

TABLE 10

Total step lengths expected under different analysis types

	Analysis Type 1		2		3	
step	exp	obs	exp	obs	exp	obs
1	9	8	8	8	8	8
2	3.83	0	3.10	0	0	0
3	0.86	0	0.63	0	0	0
4	0.11	0	0.07	0	0	0
5	0.01	0	0.00	0	0	0

(further probabilities too low for analysis to be worthwhile)

Table 10 shows the results of such a summation for the three types of analysis. Type 2 is probably the most useful set of results as it is constrained by the actual sinks in the system. Type 3 is so constrained that it reproduces that actual system exactly. The type 2 results show that probabilistically speaking there should have been about 3 two step links, whereas in fact none appeared. Again this re-emphasizes the strong structuring effect that Nairobi has on the system of cities.

The expected number of nodes tributary to node j

Adding the vectors of column sums obtained from $\bullet(\bullet)[M]^k$ from k = 1 up to some chosen value (see the last section) will yield the expected number of nodes tributary to the j^{th} node, Table 11.

TABLE 11

Total number of nodes tributary to the j^{th} node
type 2 analysis

node	1	2	3	4	5	6	7	8	9	10
exp	7.75	1.79	1.02	0.48	0.25	0.25	0.13	0.08	0.04	0.00
obs	7	0	0	0	0	1	0	0	0	0

Note: this table provides an estimate for one node at a time only.
The expected values total more than the total nodes in the system.

This is a particularly useful measure for nodes which are sinks in the actual system, as will be explained below.

Use of different analysis types for indirect flow analysis

Some care is needed in the use of the three analysis types in conjunction with the measures described in the preceding sub-sections. Type one analysis allows the evaluation of the expected nodal region size for the j^{th} node with no interruption by possible sinks lower in the hierarchy. Hence it estimates expected nodal region sizes for each node within the probabilistic constraints of •[M]. However it cannot be used as a basis for *simultaneous* comparison between several sinks and their corresponding region sizes from the actual system because the expected computations include overlapping effects, whereas the actual ones do not. Similar comments apply to the computation of the expected number of incoming k-step links and to the total number of k-step links.

Type 2 analysis specifies the actual sinks and forces them to be absorbing elements in •[M], as node 1 always is. Consequently no (probabilistic) paths are transmitted through these nodes to nodes higher up the hierarchy, Figure 3(b). For the nodes designated as sinks, absorbing states in •[M], this analysis provides estimates of nodal region sizes where the sinks are as designated, but nodes lower in the hierarchy may be sinks, passing or source nodes - sinks are already designated, and the lowest ranked node is a guaranteed source. This type of analysis may be useful when a system is almost certain to have certain centres as sinks, but where the other nodes may fluctuate between being passing nodes (receiving and sending nodal flows) and source nodes (just sending nodal flows); perhaps in circumstances where the transaction table is a sample, slight changes in certain values might affect a nodes status between 'passing' and 'source' while at the same time it may be very unlikely that it could ever change into being a sink.

Type three analysis imposes the full constraints of the actual nodal system on [M], and naturally the expected results will usually be very close to the actual ones, particularly when there are very few passing nodes - intermediate steps in the nodal hierarchy - a situation which seems to characterise actual nodal systems. However, type 3 analysis does permit the assessment of the degree to which nodes which are not sinks are distributed amongst the sinks, and this can reveal the extent to which one node, or another, is exerting an undue pull on the remaining nodes. It also shows, incidentally, that basing [M] on the column-sums of [T]* is quite reasonable. For empirical systems that I have investigated, type 3 analysis reproduces the indirect properties of the actual nodal system quite closely, but this may also result from the fact that these have been quite small systems.

Statistics for Nodal Analysis

More specific alternatives

It might be argued that if a sufficiently fine temporal timescale were taken then a transaction table could be reduced to a large number of individual events or transactions. In a given period node i might generate an outgoing transaction to node j, or it may not. For that time period the collection of transactions recorded could constitute a dichotomous matrix essentially equivalent to [F] defined above (equation 1) except that rows with zero sums might be expected. Given the existence of an external ranking scheme for the nodes, a nodal structure can be defined for the period. Consequently, a transaction table can be seen as a reflection, or a time average, of a whole series of individual events and the nodal component is a filtered subset of these events. Transforming the transaction table to a probability matrix by normalising the rows to sum to one will produce a matrix whose ij^{th} element estimates the probability that transaction ij will take place in the unit time period. This transition matrix can be used directly as the [M] matrix of the preceding sections *if* (i) a ranking and weighting of the nodes is available to reorder the matrix using [P], and then when this is done (ii) to set the upper triangular part (including the main diagonal) to zero. The ranking could be internal (based as before on column sums in the transaction table) or external, according to some criterion of geographical interest. The result will be, when [M] is treated as indicated, an analysis of the nodal structure expected during a very transient period, except that the normalisation of the rows implies that each node is active within the time unit used to construct the table. An alternative is to normalise [M] with regard to the largest row sum so that all rows except the largest sum to less than 1. This would reflect the fact that not all nodes will generate transactions within the time unit.

Regional field effects

Yet another alternative would be to specify different, and limited, field effects for different nodes and to model these in the probability matrices, [M] and ●(●)[M]. This can be done by writing down the mean information field (*sensu* Hagerstrand) for each node (a row in the probability matrix) and supplying an overall ranking of the nodes. Indirectly a similar effect could be achieved by deriving the required [M] as a transition matrix derived from some species of gravity model which estimates the $n \times n$ transaction table. A ranking of the nodes might then be based on the marginal totals of the estimated transaction table, just as in the case of an empirical tansaction table. They might, however, be supplied separately according to yet other criterion: city or regional populations, *per capita* income, or whatever might be deemed appropriate. This may be a reasonable approach for establishing the rankings of the smaller places in the system which might be particularly susceptible to variations in survey sample data, or which might have special characteristics not properly reflected in the quantitative data. For example, an administrative centre may generate transactions out of proportion to its mere population, or even its *per capita* income.

There is no need to labour these alternatives. The essential point is that the methodology can accommodate numerical inputs of any kind the investigator cares to supply, *and* can interpret.

References cited

Floyd, F.W., 1962, Algorithm 97: shortest path. *Communications ACM*, 5, 345

Kemeny, J.G. & Snell, J.L., 1960, *Finite Markov Chains*, Van Nostrand Rheinholt, New York

Nystuen, J.D. & Dacey, M.F., 1961, A graph theory interpretation of nodal regions, *Papers of the Regional Science Association*, 7, 29-42

Tinkler, K.J., 1976, On functional regions and indirect flows, *Geographical Analysis*, 8, 205-13

Tinkler, K.J., 1977, *An introduction to graph theoretical methods in Geography*, CATMOG 14, GeoAbstracts: Norwich

Tinkler, K.J., 1979, Graph Theory, *Progress in Human Geography*, 3(1), 85-116

Tinkler, K.J., 1987, *Introductory Matrix Algebra*, CATMOG 48, Geo Books: Norwich

Appendix A

1 Review of Nodal Analysis literature to 1979

2 Some selected literature on Nodal Analysis since 1979

Item 1 - A review of graph theory Nodal Analysis up to 1979

The following extract is reprinted with the permission of Edward Arnold from my essay entitled "GRAPH THEORY" which appeared in *Progress in Human Geography*, 1979, 3(1), 85-116

---- (material omitted) ----

II Graph theory--ideas from the first decade of geography

I *Nodal Analysis*

Garrison's (1960) paper started graph theory application in geography and a great many subsequent papers owe an intellectual debt to it. Huff (1960) acknowledges Garrison's influence; his matrix manipulation of variables affecting shopping behaviour was the first, and for a long time, the only non-transportation geography application. When Nystuen and Dacey (1961) introduced nodal analysis, a uniquely geographical innovation, they too brought ideas from the University of Washington. They proposed a power series expansion of the basic interaction matrix in order to account for indirect effects in the interaction system and their method had the merit of giving an automatically convergent expansion. Garrison's method of weighting the power series expansion for indirect effects was less precisely specified and may have led to non-converging expansion in later work, for example Hebert (1966), Gauthier (1968a), Harvey (1972) and Stutz (1973). A clarification of these issues was attempted by Tinkler (1974) but it is not clear that the discussion has been heeded. Cates (1978) chooses a scalar for his expansion on the basis of explorations conducted by Stutz (1973) but he does truncate the expansion at three terms: the diameter of his system. Returning to nodal analysis, the original paper has generated at least nineteen substantive applications of the method, many of them introducing variations. The commonest variation, other than the choice of analysis based on direct and/or indirect effects, has been an extension of the method by considering the n^{th} largest outflow from each node. This yields n^{th} order nodal structures. Kariel and Welling (1977), for example, carry the analysis to the fifth order. The original innovators in this extended direction were apparently Davies and Robinson (1968) whom Thorpe (1968), and see Davies (1968), takes to task on several grounds, one of which is the problem, still unrecognized or undiscussed in other applications, of how stable the nodal structures are when they are taken from sample surveys. What are the statistical properties of nodal structures, for an $n \times n$ matrix whose elements are drawn from a specified frequency distribution? For many applications, of course, the statistical reliability of the interaction data may not be called into question and indeed it is a separate problem from the structural properties of random $n \times n$ matrices.

The temporal stability of nodal structures has been examined by Clark (1973) who remarks on 'the contemporary importance, as well as the historical permanence of the underlying nodal structures' in a study of telephone calling in Wales. On the other hand Cates (1978) looks at the substantial changes in the nodal structure induced on air passenger flows as a function of the fuel crisis in 1974-5 and Clayton (1977) has examined the changing structure of interstate migration flows over several decades using nodal analysis, amongst other methods. Migayi (1966; Haggett and Chorley, 1969, 41) examined the effect on the nodal level of truncating the expansion at three different levels—as expected, higher orders of regionalization corresponded to taking more terms into the expansion. Regrettably the work was never published.

From the technical point of view a number of pleas should be made. It is not always as explicit as it should be precisely what method of analysis has been adopted—direct of indirect—and if the latter whether the matrix was scaled according to the Nystuen and Dacey method (1961) or according to some other criteria (e.g. Cates, 1978). Then again if the indirect method was used, was the power series summed directly to some convergent criterion, was it truncated at some pre-chosen value as recommended by Tinkler (1976a) or was it computed by the inverse method? Matters become more complicated when n^{th} order structures are developed. If the n^{th} order is being considered, have all the flows for orders less than n been deleted and their values deducted from the column totals, used to measure importance? The methodology on these issues is far from clear and may seriously affect attempts to replicate methodology or results.

In terms of its use as an accepted technique we find in addition to those already quoted, evidence from Soja (1968b), O'Sullivan (1968), Board *et al.* (1970), Hay and Smith (1970, 132), Waller (1970), Tolosa and Reiner (1970), Davies and Lewis (1970), Simmons (1970; 1972), Britton (1971). By the mid-1970s the method is used without extensive apologies in, for example, Riddell and Harvey (1972), Hadjifotiou (1972), Hafner (1973), Clayton (1974; 1977), Holsman (1975), Langdale (1975), Kariel and Welling (1977), and Cates (1978).

---- (material omitted) ----

V References

Note: only the references cited in the reprinted material are listed here

Board, C., Davies, R.J., and Fair, T.D.J. 1970: The structure of the South African space economy: an integrated approach. *Regional Studies*, 4, 367-92.

Britton, J.N.H. 1971: Methodology in Flow Analysis. *East Lakes Geographer*, 7, 22-36.

Cates, D.B., 1978: Short-run structural change in an active network of declining connectivity. *Professional Geographer*, 30, 9-13.

Clark, D. 1973: Communications and the urban future: a study of trunk telephone call patterns in Wales. *Regional Studies*, 7, 315-21.

Clayton, C. 1974: Communication and spatial structure. *Tijdschrift voor Economische en Sociale Geografie*, 65, 271-7.

Clayton, C. 1977: Interstate population migration process and structure in the United States, 1935–1970. *Professional Geographer*, 24, 271–7.

Davies, W.K.D. 1968: The study of nodal structure: a rejoinder. *Journal of the Town Planning Institute*, 54, 233–4.

Davies, W.K.D. and Robinson, G.W.S. 1968: The nodal structure of the Solent region. *Journal of the Town Planning Institute*, 54, 18–22.

Davies, W.K.D. and Lewis, C.R. 1970: Regional structures in Wales: two studies of connectivity. In Carter, H. and Davies, W.K.D. editors, *Urban essays in the geography of Wales*, London: Heinemann.

Garrison, W.L.,, 1960: Connectivity of the Interstate highway system. *Papers and Proceedings of the Regional Science Association*, 6, 121–37.

Gauthier, H.L. 1968a: Transportation and the growth of the Sao Paulo economy, *Journal of Regional Science*, 8, 77–94.

Hadjifotiou, N. 1972: The analysis of migration between standard metropolitan labour market areas in England and Wales. University College, London, Department of Geography, *Housing and Labour Mobility Group Working Paper*, 4.

Hafner, J.A. 1973: The spatial dynamics of rice milling and commodity flows in Central Thailand. *Journal of Tropical Geography*, 37, 30–8.

Haggett, P. and Chorley, R.J. 1969: *Network Analysis in Geography*, London: Edward Arnold.

Harvey, M.E. 1972: The identification of development regions in developing countries, *Economic Geography*, 48, 229–43.

Hay, A. and Smith, R.H.T. 1970: *Interregional trade and money flows in Nigeria 1964*. Ibadan, Nigeria: Oxford University Press.

Hebert, B. 1966: *Use of factor analysis in graph theory to identify an underlying structure of transportation network.* Ohio State University, Department of Geography (mimeo).

Holsman, A.J. 1975: Interstate interaction patterns in Australia. *Australian Geographical Studies*, 15, 41–61.

Huff, D.L., 1960: A topographical (sic) model of consumer space preferences. *Papers and Proceedings of the Regional Science Association*, 6, 159–73.

Kariel, H.G. and Welling, S.L. 1977: A nodal structure for Canadian cities using graph theory and newspaper datelines, *Canadian Geographer*, 21, 148–63.

Langdale, J.V. 1975: Nodal regional structures of New South Wales. *Australian Geographical Studies*, 13, 123–36.

Migayi, M., 1966: *Nodal regionalization of interstate highway system.* Ohio State University, Department of Geography, (mimeo).

Nystuen, J.D. and Dacey, M.F. 1961: A graph theory interpretation of nodal regions. *Papers and Proceedings of the Regional Science Association*, 7, 29–42.

O'Sullivan, P.M. 1968: Accessibility and the spatial structure of the Irish economy. *Regional Studies* 2, 195–206.

Riddell, J. B. and Harvey, M.E. 1972: The urban system in the migration process: an evaluation of step-wise migration in Sierra Leone. *Economic Geography*, 48, 270–83.

Simmons, J.W. 1970: Patterns of interaction within Ontario and Quebec. University of Toronto Centre for Urban and Community Studies, *Research Paper* 41.

Simmons, J.W. 1972: Interactions among the cities of Ontario and Quebec, In Bourne, L.S. and Mackinnon, R.D., editors, *Urban systems development in central Canada*, Toronto and Buffalo: University of Toronto Press.

Soja, E.W. 1968b: The Geography of Modernization in Kenya, *Syracuse Geographical Series*, 2, 143.

Stutz, F.P., 1973: Accessibility and the effect of scalar variations on the powered transportation matrix. *Geographical Analysis* 5, 61–6.

Thorpe, D. 1968: Some comments on the nodal structure of the Solent region. *Journal of the Town Planning Institute*, 54, 177–8.

Tinkler, K.J. 1974: On summing power series expansions of accessibility matrices by the inverse method. *Geographical Analysis*, 6, 175–8.

Tinkler, K.J. 1976a: On functional regions and indirect flows. *Geographical Analysis*, 8, 205–13.

Tolosa, H. and Reiner, T.A. 1970: The economic programming of a system of planned poles. *Economic Geography*, 46, 449–58

Waller, P.P. 1970: The delineation of a planning region: a case study from West Kenya. *East African Geographical Review*, 8, 55–60.

Item 2 - Some Nodal Analysis graph theory papers after 1979

I have found very few papers after 1979, and my intent here is to discuss two in particular, as much from the point of view of the symptons they display, as for their substantive content or interpretations.

Davies and Thompson (1980) count the Nystuen-Dacey method among the empirical methods of nodal analysis which are used when the total number of flows in interaction problems overwhelm the investigator. Other methods which were recognized to achieve simplification included factor analysis, Markov chains and cluster analysis. However all these were rejected in favour of dyadic factor analysis which can take into account multiple commodities on links, in addition to origins and destinations.

A direct use of Nystuen-Dacey analysis was made by Helleiner (1981) in a study of recreational boating on lakes, rivers and canals on the Trent-Severn waterway in Ontario. The study was perhaps unusual in the detail of the flows recorded: 156,401 separate records pertaining to 18,515 different boats and 43 locks (points of record), although only a selection of this data was used in the final study. There were other unusual aspects of the study: the flow matrix was intrinsically symmetrical because boats going into cul-de-sacs had to re-emerge. In consequence, column sums correlated with row sums with $r = 0.999$. In addition destinations with fewer than 100 boat trips recorded were dropped from the study leaving $n = 30$; these terminals being classed as *'trivial'* using Nystuen-Dacey terminology, although the definition is not precisely the same.

In the analysis that followed Helleiner differed from the standard model by looking for the largest *inflow* to a node (the largest element in its column) to define the nodal flow, rather than the largest outflow (largest element in its row). The footnote justification of this change was that since the nodal ranking was made on inflow totals (the column sums) then (logically) the nodal structure should be defined similarly. However, any node receives inflows and clearly these may be used assess its importance in the whole system (you can't be important if nobody calls you!). Nonetheless, *outflows* are needed to assess the node's connections with the rest of the system, and are hierarchically important only when they proceed up the system to more highly ranked places. Because of the high degree of symmetry in the transaction table (which was not reproduced) Helleiner argued that this would yield identical results to the original procedure. However, this is only true when the whole transaction table *is* symmetric, i.e. $T_{ij} = T_{ji}$ for all i and j, although this was probably close to being the case in his study. However, if the transaction table is *not* symmetric it will lead to differing results. Even in the small Kenyan case discussed in the main text of this monograph Helleiner's method would lead to Eldoret being connected to Nairobi, in contrast to it being a small 'upcountry' nodal centre under the standard definitions. Certainly the method of statistical analysis developed in this monograph requires that the original Nystuen-Dacey procedure be followed.

A very different approach was followed by Nader (1981). Here the intent was to develop a hierarchy of nodal regions, and for this purpose an elaborate oblique factor analysis was performed: that is to say that the factors extracted are *not* mathematically independent of each other. Nader notes that there will normally be as many nodal regions as there are factors in a rotated factor matrix, although he then admits that *'the "number-of-factors" problem has been one of the most intractable problems in factor analysis.'* Nystuen-Dacey analysis is not seriously considered until the very end, where a map is presented based on, presumably, a standard analysis. The following comment is made:

> Although the functional associations depicted by the largest flows appear to be representative of the hierarchical associations existing within the province, the Nystuen-Dacey method provides no criteria for delineating subprovincial regions.

Nader then notes that *"similar, though less extreme, polarizations have been found in other studies"* and then points to various supposed structural deficiencies in the method: (i) that important nodal centres will not necessarily be terminal points, (ii) that terminal points are not necessarily nodal centres of any consequence and the related point (iii) that terminal points may have no dependent places (*trivial* in the Nystuen-Dacey terminology). He concludes that *"in short, the Nystuen-Dacey method is simply inappropriate for delineating a hierarchy of nodal regions."*

To my prejudiced mind this is merely finding excuses for not using a simple method when a more complex one can be found; furthermore it appears to pre-judging the issue. Nader's own admission is that his Figure 8 appears to depict the functional associations in the province, a supposition re-inforced by a view of Figures 5, 6 & 7 on the previous pages which show the system deduced from the factor structure. It may be strictly true that Nystuen-Dacey analysis provides no criteria for delineating sub-provincial regions: but it is blatantly clear from his Figure 8 that such criteria are easily supplied by an enterprising research worker. The provincial 'functional associations' so well depicted are seen to be nodes which draw in nodal flows, yet which themselves are nodal to a higher ranking centre. This process can be seen to be operating over path-lengths of 3 and 4 in the Saskatchewan system. Some of the intermediate nodes have between 9 and 14 dependent places, and the lowest order ones have between 1 and 4 places. It is hard to imagine a more convincing demonstration of a nodal hierarchy! His first structural objection noted above seems to be saying that a place is only nodal if it is terminal. But that objection (or the definition on which it is based) seems to defeat the whole idea of a hierarchy in which intermediate levels act as nodal 'traps' to funnel flows up the system. His second (and the related third) objection is pre-judging the issue. It implies you already have decided what nodes are important, and why! His example would be an ideal test for the statistical apparatus developed in this monograph since it is fair comment in a graph as complex as Nader's Figure 8, (which is based on flows among 198 centres) that it *is* very hard to tell what is 'expected.' In contrast, factor models, of whatever complexion, appear to quite incapable of answering *that* question.

References

Davies, W.K.D. & Thompson, R.R. 1980: The structure of interurban connectivity: a dyadic factor analysis of Prairie commodity flows. *Regional Studies*, 14, 297-311.

Helleiner, F.M. 1981. The regionalization of a waterway: a study of recreational boat traffic. *Canadian Geographer*, 25, 60-74.

Nader, G. 1981. The delineation of a hierarchy of nodal regions by means of higher-order factor analysis. *Regional Studies*, 15, 475-492

Appendix B

FORTRAN support

1 FORTRAN 66 (old FORTRAN IV) program listing

2 Sample input file for output in monograph

3 Sample input file annotated

4 Sample output generated from 2 & 3, as used in monograph

STATISTICAL ANALYSIS OF NYSTUEN-DACEY NODAL ANALYSIS

Programmer: K.J.Tinkler Department of Geography, Brock University
St. Catharines, ONTARIO, L2S 3A1, CANADA
416-688-5550 extension 3486

{NOTE:- This is standard FORTRAN 66 (old FORTRAN IV)}

```
      DIMENSION FMT(20), MAT(30,30), IRK(30), XMAT(30,30),IC(30),IR(30)
     1,ISS(2,2),X(30),XI(30),PR(30),PC(30),EISS(30,30),PX(30),NF(30,30)
     1,IRAK(30),TITLE(20)
      READ(5,102)(TITLE(I),I=1,20)
  102 FORMAT(20A4)
  103 FORMAT(1H1,1X,20A4)
      READ(5,100)N,IEX
  100 FORMAT(2I2)
  300 FORMAT(7X,15I6,//)
      XN=N
      READ(5,102)(FMT(I),I=1,20)
      DO 1 I=1,N
    1 READ(5,FMT)(MAT(I,J),J=1,N)
      WRITE(6,103)(TITLE(I),I=1,20)
      WRITE(6,200)
  200 FORMAT(//,' NYSTUEN - DACEY ANALYSIS  NOTE: INPUT MATRIX IS LATER
     1 REORDERED BY RANK ORDER ACCORDING TO COLUMN SUMS IN INPUT MATRIX
     1 - OR BY WEIGHTS IF THIS OPTION WAS CHOSEN',///)
      WRITE(6,301)(I,I=1,N)
  301 FORMAT(4X,25I4)
      WRITE(6,203)
      DO 30 I=1,N
   30 WRITE(6,201)I,(MAT(I,J),J=1,N)
  201 FORMAT(1X,I2,2X,25I4)
      IF(IEX-1)802,801,802
  802 CONTINUE
C     ADD UP THE COLUMNS
      DO 3 J=1,N
      X(J)=0.
      DO 3 I=1,N
      IF(I-J)4,3,4
    4 X(J)=X(J)+MAT(I,J)
    3 CONTINUE
      CALL RANK(N,IRK,X)
      DO 500 I=1,N
  500 IRAK(IRK(I))=I
      IF(IEX.NE.2)GOTO 803
```

```
      DO 89 I=1,N
   89 X(I)=1000.0/(IRAK(I))
      CALL RANK(N,IRK,X)
      GOTO 803
  801 READ(5,FMT)(IC(I),I=1,N)
      DO 850 I=1,N
  850 X(I)=IC(I)
      CALL RANK(N,IRK,X)
      DO 501 I=1,N
  501 IRAK(IRK(I))=I
  803 CONTINUE
C
C     ADD UP COLUMN SUMS
C
      XS=0.
      DO 5 I=1,N
    5 XS=XS+X(I)
      CALL NODAL(N,NF,MAT,IRAK)
      CALL PERM(N,MAT,IRAK)
C
C     FIND NODAL MATRIX IN LOWER TRIANGULAR ORDERED FORM
C
C     PRINT NF AS A CHECK
C     IRK CARRIES RANK INFORMATION ABOUT ORIGINAL ORDER IN X,
C     X IS SORTED BY RANK (HIGH = 1, AND SO ON DOWNWARDS)
C
C     CONVERT X TO A PROBABILITY VECTOR
C
      DO 6 I=1,N
    6 X(I)=X(I)/XS
C
C     CONSTRUCT EXPECTED MATRIX, LOWER TRIANGULAR ONLY
C
      DO 40 I=1,N
      DO 40 J=1,N
   40 XMAT(I,J)=0.
      PR(1)=0.0
      DO 7 I=2,N
      PR(I)=0.
      DO 7 J=1,(I-1)
      XMAT(I,J)=X(J)*(1.0/(1.0-X(I)))
    7 PR(I)=PR(I)+XMAT(I,J)
C
C     XMAT IS EXPECTED MATRIX
C
C     OUTPUT BASIC OBSERVED STATISTICS
C
      IFL=0
```

```
      IINFL=0
      DO 15 I=1,N
      IR(I)=0
      IC(I)=0
      DO 15 J=1,N
      IR(I)=IR(I)+NF(I,J)
      IC(I)=IC(I)+NF(J,I)
   15 IFL=IFL+NF(I,J)
      ISKS=N-IFL
      DO 16 I=1,N
      IF(IC(I))16,16,17
   17 IINFL=IINFL+1
   16 CONTINUE
      ISRCS=N-IINFL
C
C     ISRCS = NUMBER OF SOURCES
C     ISKS  = NUMBER OF SINKS
C     COMPUTE NODE STRUCTURE TYPE FOR EACH NODE (ISS)
C
      DO 20 I=1,2
      DO 20 J=1,2
   20 ISS(I,J)=0
C
      DO 21 I=1,N
      IPT=0
      IF(IC(I))21,21,22
   22 IPT=1
   21 ISS(IR(I)+1,(IPT+1))=ISS((IR(I)+1),(IPT+1))+1
C
C     ISS IS NODE STRUCTURE MATRIX
C
C     NOW OUTPUT SO  FAR
C
      WRITE(6,202)
  303 FORMAT(9X,25I4)
  202 FORMAT(////)
  203 FORMAT(/)
      IF(IEX.EQ.1)GOTO 762
      WRITE(6,760)
  760 FORMAT(///,1X,' INPUT MATRIX REORDERED BY RANK OR COLUMN SUMS',//)
      GOTO 763
  762 WRITE(6,761)
  761 FORMAT(//,1X,' INPUT MATRIX REORDERED BY RANK OF INPUT WEIGHTS',/)
  763 CONTINUE
      WRITE(6,303)(I,I=1,N)
      WRITE(6,303)(IRK(I),I=1,N)
      WRITE(6,203)
      DO 770 I=1,N
```

```
  770 WRITE(6,222)I,IRK(I),(MAT(I,J),J=1,N)
  222 FORMAT(1X,2I3,2X,25I4)
      WRITE(6,710)
  710 FORMAT(//,' ANALYSIS IS BASED ON THIS PROBABILITY VECTOR WHICH IS
     1DERIVED FROM THE COLUMN SUMS',/,' OF THE INPUT MATRIX - - OR, IF T
     1HE OPTION WAS CHOSEN, ON AN INPUT SET OF WEIGHTS',/,9X,18F6.3,/,9X
     1,18F6.3)
      IF(IEX.EQ.1)WRITE(6,711)
  711 FORMAT(//,' IN THIS EXAMPLE INPUT WEIGHTS WERE USED',//)
      WRITE(6,203)
      WRITE(6,771)(X(I),I=1,N)
  771 FORMAT(9X,20F6.3)
      IF(IEX.EQ.2)WRITE(6,712)
  712 FORMAT(//,' THE RANK SIZED OPTION WAS USED FOR THIS ANALYSIS',//)
      WRITE(6,202)
      WRITE(6,205)
  205 FORMAT(//,' NODAL FLOWS - NODES REORDERED BY RANK',//)
      WRITE(6,206)(I,I=1,N)
      WRITE(6,206)(IRK(I),I=1,N)
  206 FORMAT(7X,30I3)
      WRITE(6,203)
      DO 32 I=1,N
   32 WRITE(6,207)I,IRK(I),(NF(I,J),J=1,N),IR(I),IC(I)
  207 FORMAT(1X,34I3)
      WRITE(6,206)(IC(I),I=1,N)
      WRITE(6,202)
      WRITE(6,202)
      WRITE(6,702)
  702 FORMAT(22X,'ROW COL')
      WRITE(6,208)
  208 FORMAT(1X,' NOTE: NODE TYPES ARE 0-I = SINK  ',/,24X,'0-0 = SINK
     1 (ISOLATED)',/,24X,'1-0 = SOURCE',/,24X,'1-I = PASSING',//)
      WRITE(6,202)
      WRITE(6,209)ISS(1,1),ISS(1,2),ISS(2,1),ISS(2,2)
  209 FORMAT(1X,' NODE TYPE FREQUENCIES',//,4X,' 0-0 = ',I4,/,5X,'0-I =
     1',I4,/,4X,' 1-0 = ',I4,/,4X,' 1-I = ',I4,///)
C
C
C     STATISTICAL TEST BASED ON XMAT
C
C
      SKEXP=0.0
      DO 440 I=1,N
  440 SKEXP=SKEXP + PR(I)
      XLINKS=SKEXP
      SKEXP=XN-SKEXP
      ILINKS=N-ISKS
      WRITE(6,350)SKEXP,ISKS,XLINKS,ILINKS
```

```
  350 FORMAT(1X,' TEST BASED ON NUMBER OF SINKS AND LINKS   ',//,'   EXP
     1ECTED SINKS = ',F6.3,'    OBSERVED SINKS = ',I3,/,'  EXPECTED LINK
     1S = ',F6.3,'    OBSERVED LINKS = ',I3,//)
C
      CHI=((SKEXP-ISKS)**2.0)/SKEXP
      SKEXP=XN-SKEXP
      XLINKS=XN-ISKS
      CHI=CHI+(((SKEXP-XLINKS)**2.0)/SKEXP)
      WRITE(6,210)CHI
  210 FORMAT(///,' CHI-SQUARED = ',F8.4'  DF = 1',//)
C
C     EXPECTED NUMBER OF NODAL FLOWS INTO A NODE
C
      DO 42 J=1,N
      PC(J)=0
      DO 42 I=1,N
   42 PC(J)=PC(J)+XMAT(I,J)
C
C
C     COMPUTATION FOR ISOLATED SINKS
C
      XISOL=0.0
      PX(N)=1.0
      DO 44 J=1,(N-1)
      XP=1.0
      DO 45 I=(J+1),N
   45 XP=XP*(1.0-XMAT(I,J))
      PX(J)=XP
       XP=XP*(1.0-PR(J))
   44 XISOL=XISOL+XP
      EISS(1,1)=XISOL
      SKEXP=XN-SKEXP
      SKOT=SKEXP-XISOL
      EISS(1,2)=SKOT
C
C     COMPUTATIONS FOR PASSING NODES CASE
C
      XXP=0.0
      DO 46 I=1,N
   46 XXP=XXP+PR(I)*(1.0-PX(I))
C
      EISS(2,2)=XXP
      EISS(2,1)=XN-(EISS(1,1)+EISS(1,2)+EISS(2,2))
      WRITE(6,202)
      WRITE(6,360)
  360 FORMAT(1X,' TEST OF STRUCTURE BASED ON THE FOUR POSSIBLE NODE TYPE
     1S',//,'  OBSERVED  EXPECTED',//)
      DO 47 I=1,2
```

```
   47 WRITE(6,216)(ISS(I,J),J=1,2),(EISS(I,J),J=1,2)
  216 FORMAT(1X,2I3,2X,2F7.3,//)
C
C     COMPUTE CHI-SQUARED
C
      CHI=0.
      DO 48 I=1,2
      DO 48 J=1,2
   48 CHI=CHI+(((ISS(I,J)-EISS(I,J))**2.0)/(EISS(I,J)))
      WRITE(6,217) CHI
  217 FORMAT(//,1X,' CHI-SQUARED = ',F7.3,' DF = 3',//)
      WRITE(6,700)
  700 FORMAT(////,'EXPECTED 1 STEP ANALYSIS')
      WRITE(6,701)
  701 FORMAT(1X,24('-'),///)
      WRITE(6,211)
  211 FORMAT(1X,' EXPECTATION THE ITH NODE HAS A NODAL FLOW VERSUS OBSER
     1VED',///)
      DO 41 I=1,N
   41 WRITE(6,212)IRK(I),PR(I),IR(I)
  212 FORMAT(1X,I3,2X,F8.4,2X,I2)
      WRITE(6,202)
      WRITE(6,213)
  213 FORMAT(1X,' EXPECTED NUMBER OF INFLOWS TO THE JTH NODE VERSUS OBSE
     1RVED',///)
      DO 43 I=1,N
   43 WRITE(6,212)IRK(I),PC(I),IC(I)
      IP=N-2
      CALL SPATH(N,NF,IRK)
      CALL TRANS(N,XMAT,IP,NF,IRK,IR,IC)
      CLOSE(6)
      STOP
      END

      SUBROUTINE TRANS(N,X,M,NF,IRK,IRR,ICC)
      DIMENSION X(30,30),X2(30,30),X3(30,30),R(30),C(30)
     1,NF(30,30),PC(30),IRR(30),ICC(30)
     1,IRK(30),IC(30),IR(30),SUM(30)
      IT=M-1
      DO 15 I=1,N
      IF(ICC(I))15,15,14
   14 ICC(I)=1
   15 CONTINUE
      WRITE(6,222)
  222 FORMAT(1H1,' TYPE 1 ANALYSIS IS ABSORBING TO THE HIGHEST RANKED NO
     1DE - NODE 1 IN THE REORDERED MATRIX',//,'  TYPE 2 ANALYSIS IS ABSO
     1RBING TO NODES DEFINED AS SINKS IN THE NODAL FLOWS MATRIX ',//,'
     1TYPE 3 ANALYSIS IS AS FOR TYPE 2 AND IN ADDITION SOURCE NODES ARE
```

```
     1CONSTRAINED SO AS NOT TO RECEIVE FLOWS')
      DO 30 KP=1,3
      DO 16 I=1,N
      DO 16 J=1,N
      GOTO(16,17,18),KP
   17 X(I,J)=X(I,J)*IRR(I)
      GOTO 16
   18 X(I,J)=X(I,J)*ICC(J)
   16 CONTINUE
      WRITE(6,220)KP
  220 FORMAT(1H1,' TYPE ',I1,' ANALYSIS')
      WRITE(6,221)
  221 FORMAT(2X,15('-'))
      CALL NORM(N,X)
      DO 40 J=1,N
      PC(J)=0.0
      DO 40 I=1,N
   40 PC(J)=PC(J)+X(I,J)
      WRITE(6,240)
  240 FORMAT(///,' EXPECTED 1 STEP ANALYSIS GIVEN THE NODAL STRUCTURE DE
     1FINED BY THE ANALYSIS TYPE',///)
      IX=1
      CALL PATHS(N,NF,IX,IC,IR,ITOT)
      T=N-1
      WRITE(6,201)IX,T,ITOT
      DO 41 I=1,N
   41 WRITE(6,203)I,IRK(I),PC(I),IC(I)
      DO 3 I=1,N
      SUM(I)=PC(I)
      DO 3 J=1,N
    3 X2(I,J)=X(I,J)
      DO 1 IT=1,(M-1)
      DO 2 I=1,N
      DO 2 J=1,N
      X3(I,J)=0.
      DO 2 K=1,N
    2 X3(I,J)=X3(I,J)+(X(I,K)*X2(K,J))
      DO 4 I=1,N
      DO 4 J=1,N
    4 X2(I,J)=X3(I,J)
      T=0.0
      DO 5 I=1,N
      R(I)=0.0
      C(I)=0.0
      DO 5 J=1,N
      R(I)=R(I)+X2(I,J)
      C(I)=C(I)+X2(J,I)
    5 T=T+X2(I,J)
```

```
      IX=IT+1
      CALL PATHS(N,NF,IX,IC,IR,ITOT)
      WRITE(6,200)IX
  200 FORMAT(////,' EXPECTED ',I2,' STEP ANALYSIS',)
      WRITE(6,210)
  210 FORMAT(1X,25('-'),///)
      WRITE(6,201)IX,T,ITOT
  201 FORMAT(1X,' TOTAL EXPECTED ',I2' STEP LINKS = ',F6.2,'  OBSERVED
     1= ',I3,///)
      WRITE(6,202)IX
  202 FORMAT(1X,' PROBABILITY THE ITH NODE HAS OUTGOING ',I2,' STEP LINK
     1 VERSUS OBSERVED',///)
      DO 6 I=1,N
    6 WRITE(6,203)I,IRK(I),R(I),IR(I)
  203 FORMAT(1X,2I4,2X,F6.2,2X,I2)
      WRITE(6,204)IX
  204 FORMAT(///,1X,' EXPECTED NO OF INCOMING ',I2,' STEP LINKS TO THE I
     1TH NODE VERSUS OBSERVED',///)
       DO 7 I=1,N
    7 WRITE(6,203)I,IRK(I),C(I),IC(I)
      DO 8 I=1,N
    8 SUM(I)=SUM(I)+C(I)
      XN=N
      TN=T/XN
      IF(TN-0.001)99,99,1
   99 ID=IX-1
      WRITE(6,211)ID
  211 FORMAT(//,' EXPECTED PROBABILITY OF A ',I2,' STEP LINK IS BELOW 0.
     101',/,' FURTHER ANALYSIS ABORTED',////)
      WRITE(6,212)
  212 FORMAT(///,1X,' EXPECTED NUMBER OF NODES TRIBUTARY TO THE JTH NODE
     1 VERSUS OBSERVED',///)
      DO 10 J=1,N
      IC(J)=0
      DO 10 I=1,N
      IF(NF(I,J))10,10,11
   11 IC(J)=IC(J)+1
   10 CONTINUE
      DO 9 I=1,N
    9 WRITE(6,203)I,IRK(I),SUM(I),IC(I)
      WRITE(6,213)
  213 FORMAT(///,1X,' NOTE - NOT ALL NODES CAN BE SINKS AT THE SAME TIME
     1 ',/,1X,' ADD 1 TO NODES WHICH ARE SINKS TO GET TOTAL NUMBER IN TH
     1E COMPONENT',///)
      GOTO 30
    1 CONTINUE
   30 CONTINUE
      END
```

```
      SUBROUTINE PATHS(N,NF,M,IC,IR,ISTPS)
      DIMENSION NF(30,30),IC(30),IR(30)
      ISTPS=0
      DO 1 I=1,N
      IC(I)=0
      IR(I)=0
      DO 2 J=1,N
      IF(NF(I,J)-M)4,3,4
    3 IR(I)=IR(I)+1
    4 IF(NF(J,I)-M)2,5,2
    5 IC(I)=IC(I)+1
    2 CONTINUE
    1 CONTINUE
      DO 6 I=1,N
    6 ISTPS=ISTPS+IR(I)
      RETURN
      END

      SUBROUTINE RANK(N,IRK,X)
      DIMENSION IRK(30),X(30)
      DO 4 I=1,N
    4 IRK(I)=I
      LIM=N-1
      DO 3 J=1,LIM
      DO 1 I=1,LIM
      M=I+1
      IF(X(I)-X(M))2,1,1
    2 XX=X(I)
      IR=IRK(I)
      X(I)=X(M)
      IRK(I)=IRK(M)
      X(M)=XX
      IRK(M)=IR
    1 CONTINUE
    3 CONTINUE
      RETURN
      END

      SUBROUTINE SPATH(N,C,IRK)
      INTEGER C(30,30),IRK(30)
C     FLOYDS SHORT PATH ALGORITHM
      DO 10 I=1,N
      DO 10 J=1,N
      IF(C(I,J))10,11,10
   11 C(I,J)=999999
   10 C(I,I)=0
C
```

```
      DO 2 K=1,N
      DO 2 I=1,N
      IF(I-K)3,2,3
    3 IF(C(I,K)-999999)7,2,7
    7 DO 4 J=1,N
      IF(J-K)5,4,5
    5 IF(C(K,J)-999999)6,4,6
    6 AD=C(I,K)+C(K,J)
      C(I,J)=AMIN1(C(I,J),AD)
    4 CONTINUE
    2 CONTINUE
      DO 20 I=1,N
      DO 20 J=1,N
      IF(C(I,J)-999999)20,22,20
   22 C(I,J)=0
   20 CONTINUE
      WRITE(6,200)
  200 FORMAT(1H1,' SHORT PATH MATRIX')
      WRITE(6,206)
  206 FORMAT(2X,17('-'))
      WRITE(6,204)(I,I=1,N)
  204 FORMAT(9X,30I3)
      WRITE(6,202)(IRK(I),I=1,N)
  202 FORMAT(9X,30I3)
      WRITE(6,207)
  207 FORMAT(/)
      DO 21 I=1,N
   21 WRITE(6,201)I,IRK(I),(C(I,J),J=1,N)
  201 FORMAT(1X,2I4,30I3)
      RETURN
      END


      SUBROUTINE NORM(N,X)
      DIMENSION X(30,30),R(30)
      DO 1 I=1,N
      R(I)=0
      DO 3 J=1,N
    3 R(I)=R(I)+X(I,J)
      IF(R(I).EQ.0)GOTO 1
      DO 2 J=1,N
    2 X(I,J)=X(I,J)/R(I)
    1 CONTINUE
      RETURN
      END


      SUBROUTINE PERM(N,IX,IRK)
```

```
      DIMENSION IX(30,30),IX2(30,30),IA(30,30),IAA(30,30),IRK(30)
      DO 1 I=1,N
      DO 1 J=1,N
      IA(I,J)=0
    1 IAA(I,J)=0
      DO 3 I=1,N
      IA(I,IRK(I))=1
    3 IAA(IRK(I),I)=1
      DO 4 I=1,N
      DO 4 J=1,N
      IX2(I,J)=0.0
      DO 4 K=1,N
    4 IX2(I,J)=IX2(I,J)+IAA(I,K)*IX(K,J)
      DO 5 I=1,N
      DO 5 J=1,N
      IX(I,J)=0.0
      DO 5 K=1,N
    5 IX(I,J)=IX(I,J)+IX2(I,K)*IA(K,J)
      RETURN
      END


      SUBROUTINE NODAL(N,NF,MAT,IRAK)
      DIMENSION NF(30,30),MAT(30,30),IRAK(30)
      DO 1 I=1,N
      DO 1 J=1,N
    1 NF(I,J)=0
      DO 8 I=1,N
      MAX=0
      DO 9 J=1,N
      IF(I-J)11,9,11
   11 IF(MAT(I,J)-MAX)9,10,10
   10 MAX=MAT(I,J)
      JJ=J
    9 CONTINUE
      IF(IRAK(I)-IRAK(JJ))8,8,12
   12 NF(IRAK(I),IRAK(JJ))=1
    8 CONTINUE
      RETURN
      END
```

2 - EXAMPLE INPUT FILE FOR KENYAN CITIES

This is the input file for the output example, which is also the example discussed in the text.

```
NODAL ANALYSIS OF KENYAN CITIES USING TELEPHONE CENSUS DATA (1967)
10 0
(20I4)
   0  14  79 189   1 172  10 166   0   6
  33   0 222   8   8 396  14 270   1   0
 121 148   0  32  35 757   3 188   2   5
 265   3  56   0   7 233   0  88   0   5
   0   3   0   2   01315   1  33   0   2
 284 230 633 1101747   0 3861153 334 577
   6   7   1   2   2 420   0 272   0   1
 109 190 191  69  301800 209   0  21  16
   4   0   2   1   2 410   1  76   0 221
   4   1   2   1  10 805   1  37 183   0
```

SAMPLE INPUT FILE FOR FORTRAN NODAL ANALYSIS PROGRAM WITH NOTES ON USE.

(The data is that used in the monograph to illustrate the method. This repeats in part the example just given in 2 above, but note the need for the final line (line NM+4) in certain cases.)

```
------------------------------------------------------------------------------
NODAL ANALYSIS OF KENYAN CITIES USING TELEPHONE CENSUS DATA (1967)
10
(20I4)
   0  14  79 189   1 172  10 166   0   6
  33   0 222   8   8 396  14 270   1   0
 121 148   0  32  35 757   3 188   2   5
 265   3  56   0   7 233   0  88   0   5
   0   3   0   2   01315   1  33   0   2
 284 230 633 1101747   0 3861153 334 577
   6   7   1   2   2 420   0 272   0   1
 109 190 191  69 301800 209   0  21  16
   4   0   2   1   2 410   1  76   0 221
   4   1   2   1  10 805   1  37 183   0
(NOTE THIS LINE FOR INTEGER WEIGHTS USING FORMAT OF LINE 3 IF LINE 2 COL 4 = 1)
------------------------------------------------------------------------------
```

NOTES ON DATA INPUT ABOVE

Line 1 :- Alphanumeric Title to label printout

Line 2 :- N in columns 1 and 2 right justified (Maximum dimension in program is currently 10). Column 3 is blank. IEX in column 4. IEX is a control variable and should be set equal to 1 in column 4 if the user is supplying an integer set of weights as a basis for estimating the expected nodal flows matrix. IEX is set equal to 2 in column 4 if the users wished to have the expected nodal flows estimated according to a simple rank-size distribution with exponent -1. If IEX is left blank (default = 0) a standard analysis based on column sums is performed.

Line 3 :- The FORTRAN format under which the input transaction matrix will be read. The program assumes that this will be an integer format, i.e. (20I4) as given here. Remeber to include the parentheses.

Lines 4 to N+3 Each row of the transaction matrix to be used for Nodal analysis. I assume here 1 line per matrix row but the format in line 3 can be used to expand this if necessary according to the usual FORTRAN format conventions. If there are M lines to one row in the transaction matrix then the last line here will be the (MN+3)th line in the data listing.

Line NM+4 If IEX in Line 2 was set to 1 then this line (or lines) is read according to the Format in Line 3, and is the integer set of weights supplied by the user. It is used by the program to estimate the expected nodal structure.

For interpretative help see the text of the monograph.

4 - OUTPUT FROM NODAL ANALYSIS PROGRAM FOR EXAMPLE IN THE MONOGRAPH

NOTE: This output is exactly what the FORTRAN program will produce for the input example provided, and is the same as the example discussed in the text. However, spacing of the output has been altered to accomodate the paging of this monograph, and to save some space. Wrapround in the titles is due to the fact that line printer output on mainframes usually allows a line longer than 80 columns, a facility that was used in the FORTRAN program.

NODAL ANALYSIS OF KENYAN CITIES USING TELEPHONE CENSUS DATA (1967)

NYSTUEN - DACEY ANALYSIS NOTE: INPUT MATRIX IS LATER REORDERED BY RANK ORDER ACCORDING TO COLUMN SUMS IN INPUT MATRIX

	1	2	3	4	5	6	7	8	9	10
1	0	14	79	189	1	172	10	166	0	6
2	33	0	222	8	8	396	14	270	1	0
3	121	148	0	32	35	757	3	188	2	5
4	265	3	56	0	7	233	0	88	0	5
5	0	3	0	2	0	1315	1	33	0	2
6	284	230	633	110	1747	0	386	1153	334	577
7	6	7	1	2	2	420	0	272	0	1
8	109	190	191	69	30	1800	209	0	21	16
9	4	0	2	1	2	410	1	76	0	221
10	4	1	2	1	10	805	1	37	183	0

INPUT MATRIX REORDERED BY RANK OR COLUMN SUMS

		1	2	3	4	5	6	7	8	9	10
		6	8	5	3	10	1	7	2	9	4
1	6	0	1153	1747	633	577	284	386	230	334	110
2	8	1800	0	30	191	16	109	209	190	21	69
3	5	1315	33	0	0	2	0	1	3	0	2
4	3	757	188	35	0	5	121	3	148	2	32
5	10	805	37	10	2	0	4	1	1	183	1
6	1	172	166	1	79	6	0	10	14	0	189
7	7	420	272	2	1	1	6	0	7	0	2
8	2	396	270	8	222	0	33	14	0	1	8
9	9	410	76	2	2	221	4	1	0	0	1
10	4	233	88	7	56	5	265	0	3	0	0

ANALYSIS IS BASED ON THIS PROBABILITY VECTOR WHICH IS DERIVED FROM THE COLUMN SUMS
OF THE INPUT MATRIX - - OR, IF THE OPTION WAS CHOSEN, ON AN INPUT SET OF WEIGHTS

0.408 0.148 0.119 0.077 0.054 0.053 0.040 0.039 0.035 0.027

NODAL FLOWS - NODES REORDERED BY RANK

		1	2	3	4	5	6	7	8	9	10		
		6	8	5	3	10	1	7	2	9	4		
1	6	0	0	0	0	0	0	0	0	0	0	0	7
2	8	1	0	0	0	0	0	0	0	0	0	1	0
3	5	1	0	0	0	0	0	0	0	0	0	1	0
4	3	1	0	0	0	0	0	0	0	0	0	1	0
5	10	1	0	0	0	0	0	0	0	0	0	1	0
6	1	0	0	0	0	0	0	0	0	0	0	0	1
7	7	1	0	0	0	0	0	0	0	0	0	1	0
8	2	1	0	0	0	0	0	0	0	0	0	1	0
9	9	1	0	0	0	0	0	0	0	0	0	1	0
10	4	0	0	0	0	0	1	0	0	0	0	1	0
		7	0	0	0	0	1	0	0	0	0		

ROW COL
NOTE: NODE TYPE S ARE 0-I = SINK
0-0 = SINK (ISOLATED)
1-0 = SOURCE
1-I = PASSING

NODE TYPE FREQUENCIES

0-0 = 0
0-I = 2
1-0 = 8
1-I = 0

TEST BASED ON NUMBER OF SINKS AND LINKS

EXPECTED SINKS = 2.709 OBSERVED SINKS = 2
EXPECTED LINKS = 7.291 OBSERVED LINKS = 8

CHI-SQUARED = 0.2548 DF = 1

TEST OF STRUCTURE BASED ON THE FOUR POSSIBLE NODE TYPES

OBSERVED		EXPECTED	
0	2	0.897	1.812
8	0	5.674	1.616

CHI-SQUARED = 3.486 DF = 3

EXPECTED 1 STEP ANALYSIS

EXPECTATION THE ITH NODE HAS A NODAL FLOW VERSUS OBSERVED

6	0.0000	0
8	0.4789	1
5	0.6311	1
3	0.7312	1
10	0.7947	1
1	0.8512	0
7	0.8954	1
2	0.9357	1
9	0.9722	1
4	1.0000	1

EXPECTED NUMBER OF INFLOWS TO THE JTH NODE VERSUS OBSERVED

6	3.9395	7
8	1.2524	0
5	0.8752	0
3	0.4804	0
10	0.2804	0
1	0.2216	1
7	0.1255	0
2	0.0796	0
9	0.0360	0
4	0.0000	0

SHORT PATH MATRIX

		1	2	3	4	5	6	7	8	9	10
		6	8	5	3	10	1	7	2	9	4
1	6	0	0	0	0	0	0	0	0	0	0
2	8	1	0	0	0	0	0	0	0	0	0
3	5	1	0	0	0	0	0	0	0	0	0

4	3	1	0	0	0	0	0	0	0	0	0
5	10	1	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0
7	7	1	0	0	0	0	0	0	0	0	0
8	2	1	0	0	0	0	0	0	0	0	0
9	9	1	0	0	0	0	0	0	0	0	0
10	4	0	0	0	0	0	1	0	0	0	0

TYPE 1 ANALYSIS IS ABSORBING TO THE HIGHEST RANKED NODE - NODE 1 IN THE REORDERED MATRIX

TYPE 2 ANALYSIS IS ABSORBING TO NODES DEFINED AS SINKS IN THE NODAL FLOWS MATRIX

TYPE 3 ANALYSIS IS AS FOR TYPE 2 AND IN ADDITION SOURCE NODES ARE CONSTRAINED SO AS NOT TO RECEIVE FLOWS

TYPE 1 ANALYSIS

EXPECTED 1 STEP ANALYSIS GIVEN THE NODAL STRUCTURE DEFINED BY THE ANALYSIS TYPE

TOTAL EXPECTED 1 STEP LINKS = 9.00 OBSERVED = 8

1	6	5.17	7
2	8	1.51	0
3	5	1.00	0
4	3	0.53	0
5	10	0.30	0
6	1	0.23	1
7	7	0.13	0
8	2	0.08	0
9	9	0.04	0
10	4	0.00	0

EXPECTED 2 STEP ANALYSIS

TOTAL EXPECTED 2 STEP LINKS = 3.83 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 2 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.27	0

4	3	0.40	0
5	10	0.46	0
6	1	0.49	0
7	7	0.52	0
8	2	0.55	0
9	9	0.56	0
10	4	0.58	0

EXPECTED NO OF INCOMING 2 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	2.97	0
2	8	0.53	0
3	5	0.21	0
4	3	0.07	0
5	10	0.03	0
6	1	0.01	0
7	7	0.01	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 3 STEP ANALYSIS

TOTAL EXPECTED 3 STEP LINKS = 0.86 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 3 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.05	0
5	10	0.08	0
6	1	0.11	0
7	7	0.13	0
8	2	0.15	0
9	9	0.17	0
10	4	0.18	0

EXPECTED NO OF INCOMING 3 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.75	0
2	8	0.08	0
3	5	0.02	0
4	3	0.01	0
5	10	0.00	0

6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 4 STEP ANALYSIS

TOTAL EXPECTED 4 STEP LINKS = 0.11 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 4 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.01	0
7	7	0.02	0
8	2	0.02	0
9	9	0.03	0
10	4	0.03	0

EXPECTED NO OF INCOMING 4 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.10	0
2	8	0.01	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 5 STEP ANALYSIS

TOTAL EXPECTED 5 STEP LINKS = 0.01 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 5 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0

3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED NO OF INCOMING 5 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.01	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED PROBABILITY OF A 4 STEP LINK IS BELOW 0.01
FURTHER ANALYSIS ABORTED

EXPECTED NUMBER OF NODES TRIBUTARY TO THE JTH NODE VERSUS OBSERVED

1	6	9.00	7
2	8	2.13	0
3	5	1.24	0
4	3	0.61	0
5	10	0.33	0
6	1	0.25	1
7	7	0.13	0
8	2	0.08	0
9	9	0.04	0
10	4	0.00	0

NOTE - NOT ALL NODES CAN BE SINKS AT THE SAME TIME
ADD 1 TO NODES WHICH ARE SINKS TO GET TOTAL NUMBER IN THE COMPONENT

TYPE 2 ANALYSIS

EXPECTED 1 STEP ANALYSIS GIVEN THE NODAL STRUCTURE DEFINED BY THE ANALYSIS TYPE

TOTAL EXPECTED 1 STEP LINKS = 9.00 OBSERVED = 8

1	6	4.67	7
2	8	1.33	0
3	5	0.86	0
4	3	0.44	0
5	10	0.24	0
6	1	0.23	1
7	7	0.13	0
8	2	0.08	0
9	9	0.04	0
10	4	0.00	0

EXPECTED 2 STEP ANALYSIS

TOTAL EXPECTED 2 STEP LINKS = 3.10 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 2 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.27	0
4	3	0.40	0
5	10	0.46	0
6	1	0.00	0
7	7	0.46	0
8	2	0.49	0
9	9	0.51	0
10	4	0.53	0

EXPECTED NO OF INCOMING 2 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	2.46	0
2	8	0.41	0
3	5	0.15	0
4	3	0.05	0
5	10	0.02	0
6	1	0.01	0
7	7	0.01	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 3 STEP ANALYSIS

TOTAL EXPECTED 3 STEP LINKS = 0.63 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 3 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.05	0
5	10	0.08	0
6	1	0.00	0
7	7	0.10	0
8	2	0.12	0
9	9	0.13	0
10	4	0.15	0

EXPECTED NO OF INCOMING 3 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.56	0
2	8	0.05	0
3	5	0.01	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 4 STEP ANALYSIS

TOTAL EXPECTED 4 STEP LINKS = 0.07 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 4 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.01	0
8	2	0.01	0
9	9	0.02	0
10	4	0.02	0

EXPECTED NO OF INCOMING 4 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.06	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 5 STEP ANALYSIS

TOTAL EXPECTED 5 STEP LINKS = 0.00 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 5 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED NO OF INCOMING 5 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED PROBABILITY OF A 4 STEP LINK IS BELOW 0.01

FURTHER ANALYSIS ABORTED

EXPECTED NUMBER OF NODES TRIBUTARY TO THE JTH NODE VERSUS OBSERVED

1	6	7.75	7
2	8	1.79	0
3	5	1.02	0
4	3	0.48	0
5	10	0.25	0
6	1	0.25	1
7	7	0.13	0
8	2	0.08	0
9	9	0.04	0
10	4	0.00	0

NOTE - NOT ALL NODES CAN BE SINKS AT THE SAME TIME
ADD 1 TO NODES WHICH ARE SINKS TO GET TOTAL NUMBER IN THE COMPONENT

TYPE 3 ANALYSIS

EXPECTED 1 STEP ANALYSIS GIVEN THE NODAL STRUCTURE DEFINED BY THE ANALYSIS TYPE

TOTAL EXPECTED 1 STEP LINKS = 9.00 OBSERVED = 8

1	6	7.54	7
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.46	1
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED 2 STEP ANALYSIS

TOTAL EXPECTED 2 STEP LINKS = 0.00 OBSERVED = 0

PROBABILITY THE ITH NODE HAS OUTGOING 2 STEP LINK VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0

4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED NO OF INCOMING 2 STEP LINKS TO THE ITH NODE VERSUS OBSERVED

1	6	0.00	0
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.00	0
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

EXPECTED PROBABILITY OF A 1 STEP LINK IS BELOW 0.01
FURTHER ANALYSIS ABORTED

EXPECTED NUMBER OF NODES TRIBUTARY TO THE JTH NODE VERSUS OBSERVED

1	6	7.54	7
2	8	0.00	0
3	5	0.00	0
4	3	0.00	0
5	10	0.00	0
6	1	0.46	1
7	7	0.00	0
8	2	0.00	0
9	9	0.00	0
10	4	0.00	0

NOTE - NOT ALL NODES CAN BE SINKS AT THE SAME TIME
ADD 1 TO NODES WHICH ARE SINKS TO GET TOTAL NUMBER IN THE COMPONENT

APPENDIX C

"A graph theory interpretation of nodal regions"

by John D. Nystuen and Michael F. Dacey

This paper appeared originally in the *Papers and Proceedings of the Regional Science Association*, Volume 7, 1961, pp. 29–42. It is reprinted here with the permission of that journal which retains all rights. It also appears here with the permission of both authors.

A GRAPH THEORY INTERPRETATON OF NODAL REGIONS

By John D. Nystuen and Michael F. Dacey

The authors are, respectively, Assistant Professor of Geography in the University of Michigan and Assistant Professor of Regional Science in the Uuiversity of Pennsylvania. This study was formulated and the computations were obtained while both authors were at the University of Washington, Seattle. Professor Nystuen has been the primary contributor to the present statement. The responsibility for this statement is, of course, shared jointly.

THE PURPOSE OF THIS PAPER is to describe a procedure for ordering and grouping cities by the magnitude and direction of the flows of goods, people, and communications between them. Current theories of nodal regions and central place hierarchies provide the bases for the recognition of region-wide organization of cities into networks. These two theories were developed by students who recognized that the direction and magnitude of flows associated with social processes are indicators of spatial order in the regional structure of urban society. Whether the flow is local and to the city's hinterland, or regional and to the rank ordering of cities, the notion of central or nodal point is dependent upon the levels of strongest associations within the total flow.[1]

The present problem is to develop a method capable of quantifying the degree of association between city pairs in a manner that allows identification of the networks of strongest association. These associations may be in terms of interactions that occur directly between two cities, or indirectly through one or more intermediary cities. The magnitude of the combined direct and indirect associations is measured by an index that is related to certain concepts of graph theory. This index is used to identify the degree of contact between city pairs and it provides a quantitative basis for grouping cities. The resulting subgroups of cities are analogous to nodal regions. When each city in a study region is assigned to a subgroup, it is possible to specify the rank ordering of cities and to evaluate the functional relations of the nodal hierarchy.

In this paper, pertinent geograrphic and graph theoretic concepts are discussed and are then used as a basis for deriving the method of isolating nodal regions. While this method is illustrated by the use of intercity telephone calls in Washington state, the techniques are quite general and may be adapted to

[1] Berry, B.J.L. and W.L. Garrison, "A Note on Central Place Theory and the Range of a Good," *Economic Geography*, Vol. 34 (1958), pp. 304-311. Ullman, E.L., "A Theory for Location of Cities," *American Journal of Socialogy*, Vol. 46 (1941), pp. 853-864. Whittlesey D., "The Regional Concept and the Regional Method," *American Geography: Inventory and Prospect*, (P.C. James and C.F. Jones, eds.) Syracuse University, Syracuse, 1954.

PAPERS AND PROCEEDINGS OF THE REGIONAL SCIENCE ASSOCIATION, VOLUME 7, 1961

many types of phenomena. A particular phenomenon is suitable for this type of analysis when it may be viewed as a relationship or flow that links objects that are properly mapped as points. In the present illustration, cities are conceptualized as punctiform elements in a telephone network. Other suitable areas of application include the flow of information or material products between business firms in a metropolitan area, the flow of mail or freight between cities in a region, the interpersonal relations between the inhabitants of a city or the political structure that connects federal, state and local governments.

Relationship to Existing Theory

Cities may be viewed as nuclei of specialized activities which are spatially concentrated and functionally associated. Each activity has its own set of associations outside the city. To account for the many different external connections of each specialization, general statements concerning urban associations must be multi-dimensional. Accordingly, urban hinterlands are normally defined by establishing a boundary from a composite of the spatial range of several central place functions, such as the trade area of the local newspaper, the extent of wholesale drug distribution, bus passenger volumes, governmental jurisdiction, and similar indices of central place functions.

Long-distance telephone communications may be considered a single index of this multi-dimensional association among cities. A grouping of cities on the basis of telephone data defines only a network of telephone traffic centers. The validity of interpreting these telephone traffic centers as an accurate indicator of multifunctional associations depends upon a correspondence of the hinterlands which are developed with those obtained from studies which evaluate many types of contacts. The authors are willing to accept that telephone flows are one of the best single indices of all functional contacts. It has an advantage over the use of a series of indices because it obviates weighing the individual contributions of the several indices.[2]

The Nodal Region or Hinterland

Nodal regions are defined by evaluating the external contacts of small areal units. Each of these areal units is assigned to that place with which it has the dominant association. Usually, this will be a nearby city, and this city is defined as the central place or nodal point for the unit areas oriented to it. The aggregation of these unit areas, in turn, is called the nodal region.

This does not deny the existence of other flows or associations to and from each areal unit. Such flows do exist so that each areal unit is connected to many other cities. Newspaper circulation, for example, may be dominated by the local daily while the nearest metropolitan paper may also be well represented, and *The New York Times* may find its way into a few homes in the area. Also, many sporatic contacts with former hometown papers may be present.

[2] Hammer, C. and F.C. Iklé, "Intercity Telephone and Airline Traffic Related to Distance and the 'Propensity to Interact'," *Sociometry*, Vol. 20 (1957), pp. 306-316. Harris, C.D., *Salt Lake City: A Regional Capital*, University of Chicago, Chicago, 1940. Ullman, E.L., *Mobile: Industrial Seaport and Trade Center*, University of Chicago, Chicago, 1940.

Nevertheless, the "dominant association" remains the critical concept in defining a nodal structure. The remaining non-dominant associations are not used, even though the magnitudes of some of these associations may be relatively large.

The Hierarchy of Cities

The nodal rgion describes the relationship between the hinterland, which is areal, and the central or nodal city, which is punctiform. Clearly, there is no loss of generality by considering only paired contacts between points. In the hinterland concept, the areal units may be abstracted to the level of points so that the association is in terms of many points being linked to a single central point.[3]

The hinterland of a major metropolitan center, such as Chicago, may encompass a large region and incorporate many of the region's functions. The strongest of the flows between Chicago and its hinterland are point to point associations of the cities within the region. At this scale, the relationship between nodal regions and the hierarchy of central places becomes clear. The major hinterland of Chicago is defined by its dominant association with many smaller metropolises. Each of these centers, in turn, is the focus of association from other, smaller centers within its immediate vicinity. These associations incorporate lower-order functions than those establishing direct associations to Chicago. In this fashion city regions are nested together, intimately dependent upon the range of the functions which define the associations at each level.

A hierarchy of cities of this type may be reduced to an abstract network of points and lines. The points represent the cities while the lines represent the functional associations. Though a myriad of lines exists in the network, there is present a basic structure of strongest associations which creates the nested nodal regions and the hierarchy of cities. Both the direct and indirect associations are important in these intercity structures. In terms of the direct associations, for example, a wholesale establishment may receive orders directly and ship directly to some points within the system. Alternately, the associations are indirect when the orders are accumulated at various levels of the hierarchy and proceed upward to the regional headquarters. In the same manner, the outbound shipments from the central city proceed down the ranks to intermediary levels through middlemen, rather than directly to every point in the region.

Many associations are of this indirect type. For instance, political control moves up and down the ranks, rather than through direct communication between the national party leaders and the ward leaders. Most commodities are assembled and distributed through a hierarchical structure within the organization. This results, in part, from the economies of moving large lots over long distances and, in part, from the better control it affords over the operation. In evaluating the entire fabric of urban society, it is evident that subtle, indirect influences and associations are frequently exerted by one location on another. A system of analysis which accounts for both the direct and the indirect as-

[3] For example see: Isard, W. and D.J. Ostroff., "General Interregional Equilibrium," *Journal of Regional Science*, Vol. 2 (1960) pp. 67-74.

sociations between cities is appropriate.

In summary, the nodal region is defined on the basis of the single strongest flow emanating from or moving to each of the unit areas in the vicinity of a central place. The region is delimited by the aggregation of these individual elements. The hierarchy of central places is determined by the aggregation of the smallest central places which are dependent upon a single, larger center for the functions they lack. This nesting of cities defines the organization of networks of cities and the position of each city within the network. Such nesting depends upon the available bundle of functions and the relative dominance of bundles.

In this study we start with the cities and towns of a large area. Then, the structure of association among the cities is specified by assigning each city to one of several subgroups. By considering the system as a set of points and lines, where the lines represent the association between points, certain theorems of linear graphs become available for the analysis of the functional association of cities within an area.

A GEOGRAPHICAL APPLICATION OF SOME GRAPH THEORY CONCEPTS

Graph theory is a mathematics of relations. By specifying certain properties of the relations between cities and accepting the point-line abstraction of graph theory, certain theorems become available for analyzing intercity flows.[4] Consider the cities in a region as a set of points. Consider, also, a line joining a pair of points whenever there exists a certain flow between the cities they represent. The finite collection of points and lines, where each line contains exactly two points, is a linear graph of the relations established by the flows.

Some Characteristics of Linear Graphs

A point is called *adjacent* to another point if it is connected to it by a line. The network of lines is the only information contained within the graph. Scalar distance and direction, the most striking aspects of geographical maps, are not defined for a graph. If the relationship is of equal value for every connected pair, the graph is a *binary graph*. Most graph theory relates to this type of construction which simply indicates whether a line (a relation) exists or does not exist between any pair of points. The connections, however, may be considered to have intensity. *Intensity* is displayed on the graph by assigning a value to the lines.

Orientation of a relation between two points is displayed on the graph by an arrowhead $a \rightarrow b$, and read "a is related to b." A graph which specified

[4] Some general statements of graph theory are: König, D., *Theorie der Endlichen und Unenlichen Graphen*. Leipzig, 1936 (reprinted by Chelsea Publishing Co., New York, 1950); Berge, C., *Theorie des Graphes et Ses Applications*. Dunod, Paris, 1958; Harary, F., "Unsolved Problems in the Enumeration of Graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, Vol. 5 series A (1960) pp. 63-95;—, "Some Historical and Intuitive Aspects of Graph Theory," *Siam Review* Vol. 2 (April 1960) pp. 123-131. The utility of graph theory for geographic analysis has been demonstrated by Garrison, W.L., "Connectivity of the Interstate Highway System," *Papers and Proceeding of The Regional Science Association*, Vol. 6 (1960) pp. 121-137.

orientation is called a *directed graph* or *digraph*. The relationship between two points on a directed graph need not be symmetrical and, when intensity of the connection is defined, the intensity may be different for each direction.

A path from the points a to e is a collection of points and lines of the form, $a, a \rightarrow b, b, b \rightarrow c, \cdots, d, d \rightarrow e, e,$ where the points $a, b, \cdots, e,$ are distinct. A *sequence* is a collection of points and lines from a to e in which the intermediate points need not be distinct. A graph is *weakly connected* if there exists a path between each pair of points, disregarding orientation. The points in a *component* of a graph are weakly connected and are not connected to any other points in the graph. The *degree* of a point is the number of points to which it is adjacent. In a directed graph a point has an out-degree and an in-degree depending on the orientation of the lines incident to it.

Matrix Notation

For every linear graph there is an *adjacency matrix* which completely describes the graph, and *vice versa*. The matrix notation is convenient for arithmetic manipulation. Every point in a graph is represented by a row and a column of the matrix. The element, x_{ij}, of the adjacency matrix takes the value of the line; if it exists, between the points i and j; if the line does not exist, the value of the x_{ij} is O.

The diagonal elements, x_{ii}, of the adjacency matrix represent the relation of each point to itself. This relationship may or may not be defined. When it is not defined, all elements of the main diagonal are, by convention, put equal to zero.

Properties of the Dominant Relations Between Cities

The geographic theory reviewed above suggests that within the myraid relations existing between cities, the network of largest flows will be the ones outlining the skeleton of the urban organization in the entire region. The term "largest" implies an oriented relation because a flow between a pair of cities may be the largest in terms of one city but not necessarily in terms of the other city. The relation "largest flow" may have various definitions, such as the largest out-flow, in-flow, or total flow. The present example uses the number of out-going intercity telephone messages from each city to every other city in the study area. It is possible to construct a directed graph of these relations. Using the principle of dominant association, a single out-directed line is assumed to be associated with each point. When number of telephone messages are used to measure intensity of intercity associations, this assumption is easily accepted because for any city the largest volume to any one city is typically several times greater than the next largest message flow. An assumption of this type is tenable only for intercity relations which may be ranked or have a unique, largest interaction. In other situations, nodal region is most likely an inappropriate concept.

The collection of largest flow lines between city pairs defines a network of orientation among the points. Where each point has a largest flow, that largest flow may be found by simple inspection of a matrix of flows between all pairs, and it is the maximum element in each row when the matrix displays number of messages from the row city to the column city. The present intention is to use this notion of largest flow to aggregate cities associated with a central place. The resulting aggregation is said to be composed of the "subordinates" of the central city. The problem is the recognition of a "central city." In order to establish a "dominate center" three additional properties of the "largest flow" relation are now identified.

One property states that a city is "independent" if its "largest flow" is to a "smaller city." A small city remote from large metropolitan centers may display this type of independence because its largest flow is to an even smaller, nearby city. Conversely, in the same region a large satellite city closely associated with a metropolitan center does not have this independence because its largest flow is to the metropolis. So, to identify independent cities a measure of size is required. Size may be externally assigned, e.g., by population of each city; or it may be internally assigned, e.g., by the total volume of messages to or from all cities in the region. In the example below size is assigned in accordance with the total in-message flow from all cities in the study region. This value is the column total of the matrix of flows between all pairs of cities. In these terms, an "independent" or "central city" is defined as one whose largest flow is to a smaller city: A subordinate city is a city whose largest flow is to a larger city. This assumes no ambiguities arise to obscure the dominate (largest) city of a pair. This occurs when largest flows are reflexive, that is, two cities whose largest out-connections are to each other.

A second property is transitivity. This property implies that if a city a is subordinate to city b and b is subordinate to c, then a is subordinate to c.

A third property stipulates that a city is not a subordinate of any of its subordinates. A graph showing this relation is called *acyclic*. It is easily seen that an acyclic graph contains a *hierarchy*.

Two Theorems

The largest flow from every subordinate city is called the *nodal flow*. These flows form the *nodal structure* of the region and (for the particular relation under study) this skeleton displays the functional association of the cities in a region. This structure is analogous the nodal region and contains a hierarchy of centers. It is important to recognize that in this nodal structure the out-contact from at least one point is zero. This particular case is called a *terminal point* and in terms of an urban structure, this type of point is interpreted as a central city.

The following statements are useful deductions concerning the graph of a nodal structure. Figure 1 illustrates the resulting concepts.

(1) The components of a nodal structure partition the set of cities.

Proof. Each city is represented by a point in a graph of a nodal structure. Each point is weakly connected to every point in its component and to no point not in its component. A trivial component is an isolated point. Each point is, therefore, assigned to one and only one component. Such an assignment of a set is a partitioning.

(2) Each component of a nodal structure has a unique central city (terminal

MATRIX OF NUMBER OF MESSAGES BETWEEN CITY-PAIRS

TO CITY

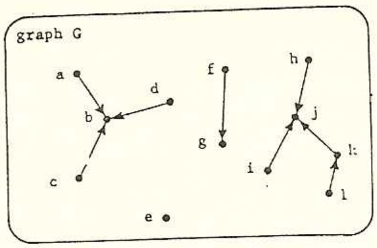
FROM CITY	a	b	c	d	e	f	g	h	i	j	k	l
a	0	(75)	15	20	28	2	3	2	1	20	1	0
b*	69	0	45	50	58	12	20	3	6	35	4	2
c	5	(51)	0	12	40	0	6	1	3	15	0	1
d	19	(67)	14	0	30	7	6	2	11	18	5	1
e*	7	40	48	26	0	7	10	2	37	39	12	6
f	1	6	1	1	10	0	(27)	1	3	4	2	0
g*	2	16	3	3	13	31	0	3	18	8	3	1
h	0	4	0	1	3	3	6	0	12	(38)	4	0
i	2	28	3	6	43	4	16	12	0	(98)	13	1
j*	7	40	10	8	40	5	17	34	98	0	35	12
k	1	8	2	1	18	0	6	5	12	(30)	0	15
l	0	2	0	0	7	0	1	0	1	6	(12)	0
Column Total	113	337	141	128	290	71	118	65	202	311	91	39

Largest flow circled. Largest flow determined by the number of out-going messages.

*Largest flow from these cities is to a "smaller" city where "size" is determined by the column totals.

GRAPH OF THE NODAL STRUCTURE BETWEEN CITIES

Graph of a,b,..., l cities in Region G.



Adjacency Matrix of Graph G

	a	b	c	d	e	f	g	h	i	j	k	l	
a		1											
b													*
c		1											
d		1											
e													**
f							1						
g													*
h										1			
i										1			
j													*
k										1			
l											1		

Blank spaces represent zero elements.

*Terminal point. **Trivial terminal point.

FIGURE 1. Graph of a Nodal Structure in a Region. (Hypothetical)

point).

Proof. Every path has at least one subordinate point and also a point to which all points on the path are subordinate by the transitive property of the relation. If this point is subordinate it must be adjacent to a point not on the path because the relation is acyclic. Upon extending the path, an end point with zero out-degree will be found in the component because at least one point is in the component and is not subordinate.[5] This end point is the only terminal

[5] Here, only finite graphs are considered.

point to which all points on the related paths are subordinate because no branching occurs on any path (every subordinate point has an out-degree of one).

Now assume distinct paths in a component and extend the paths to their terminal points. Any other point connected to a point in one extended path has no other connection because each point has an out-degree no greater than one. If their terminal points were distinct, the subgraphs associated with distinct paths by the method described would not be connected in any way. This contradicts the fact that all elements in a component are weakly connected. Therefore, each extended path must have the same terminal point.

Interpretation of the Nodal Structure

The nodal structure may be used to distinguish groups of cities that have maximum direct linkages and the rank order of these cities may be calculated. The hinterland of the central city may also be determined by mapping the cities in a nodal structure and then drawing a line just beyond the cities which are most distant from the central city. In accordance with existing theory, the hinterland or nodal region contains the area in which the maximum association or flow is toward the central or nodal city. In addition, this plotting shows the hierarchy over which the central city is dominant.

An Extention of the Theory to Indirect Associations

The operations and structure that has been described evaluates only direct contacts between city-pairs. It does not incorporate the indirect associations, and these, conceivably, could be very influential in determining functional associations. Admittedly, direct contacts should receive the greater weight, but some evaluation of the indirect channels between city-pairs would seem appropriate because of the indirect associations which occur within a hierarchy. Indirect associations may be evaluated by using matrix manipulations to adjust the nodal structure. It is postulated that the increment of indirect association or influence decreases with increases in the length of the channel.

Power Series of the Adjacency Matrix

The first step in accounting for indirect influence is to adjust the raw data matrix so that the direct association between each city-pair is some proportion of the total association of the largest center in the area. This is accomplished by obtaining the maximum column total of the adjacency matrix ($\max \sum_j x_{ij}$) and dividing every x_{ij} element by this summation. Put: $y_{ij} = x_{ij}/\max \sum_j x_{ij}$. The following inequalities result for a graph of n points:

$$0 \leqq y_{ij} < 1 \qquad (i, j = 1, 2, \cdots, n) \tag{1}$$

$$0 < \sum_j y_{ij} \leqq 1 \qquad (j = 1, 2, \cdots, n)\,. \tag{2}$$

The maximum column total equals *1*.

The linear graph corresponding to this adjacency matrix has the appropriate, positive, decimal loading. Let the adjacency matrix be called Y. In terms of linear graphs, the power expansions of Y have interesting interpretations. The matrix Y^2, which is obtained by $Y \cdot Y$ under usual matrix multiplication, describes a graph when all sequences have a length of *2*. The length of a sequence is the number of lines it contains. Further, the loading of the lines of each sequence of length *2* are obtained by multiplication. Since the initial loadings are decimal values, an attenuated value is associated with a contact that proceeds from point i to j through a sequence of length *2*. The sum of all such two-step sequences from i to j is the value of all possible indirect contacts of length *2*.

This assertion may be demonstrated as true by considering the meaning of the summation:

$$a_{ij} = \sum_k y_{ik} y_{kj} \qquad (k = 1, 2, \cdots, n) \tag{3}$$

and where a_{ij} is an element in Y^2.

The y_{ik} is the loading on the line from point i to point k in the graph and the y_{kj} has the same meaning for the link from k to j. The only terms which enter the summation are those where a sequence of length *2* exists. When a link from or to the kth point does not exist, the whole term is zero. The a_{ij} is the total value of all sequences of length *2*. In a similar manner it may be shown that the elements of Y^3 specify the attenuated value of all sequences of length *3*, and so on. The meaning of the following summation is clear:

$$B = Y + Y^2 + Y^3 + \cdots + Y^n + \cdots \tag{4}$$

The element, b_{ij}, of B represents the total direct and indirect influence from i to j.

Some examples may be useful. Given the cities $a, b, \cdots, n$, a typical sequence from a to e might be $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e$. Imagine an activity in city a as having influence on a respondent in b, this b in turn contacts a respondent in c, and continuing until a small response in e is affected. The probability of such a chain of occurrence depends, in part, on the magnitude of the flows in every link of the sequence. In general, the longer the sequence, the more remote is the probability of a response and when a response occurs it is less intense.

Alternatively, the flow of influence may be re-channeled through the same city more than once. For example, a sequence may have the form $a \rightarrow b \rightarrow a \rightarrow e$. All such summations are included in the matrix B.

The summation of the power expansion of Y is not demonstrated to be the correct form of the attenuation of flows in a sequence. It is extremely doubtful that the matrix B is the most appropriate measurement of the total direct and indirect influences. It is essentially a measure of chance indirect contact. The distribution of actual indirect association is very likely not at all random but rather concentrated in certain flow channels, in which case the matrix B would be an underestimate of indirect influence. It does, however, have a greater appeal than the matrix Y which incorporates only the direct influences. The choices of the particular power expansion is dictated by the ease of its computation. Several other methods may also be appropriate.[6]

Computation of the Power Series of the Adjacency Matrix

A convenient method of computing the matrix B is to use the following identity:

$$(1 - Y)^{-1} = 1 + Y + Y^2 + \cdots + Y^n + \cdots \tag{5}$$

and then:

$$B = (1 - Y)^{-1} - 1 \, , \tag{6}$$

where the 1 is the identity matrix. The inverse, $(1 - Y)^{-1}$, is known to exist if the inequalities (1) and (2) hold.

The Nodal Structure of Matrix B

The nodal structure of matrix B is established by isolating the network of largest flows in the same manner as was described for the direct associations. Because the associations enumerated in matrix B are adjusted for both direct and indirect flows, it is expected that a more reasonable structure is obtained.

AN EXAMPLE

Washington State was chosen as the study area. The utility of the nodal structure concept is evaluated by choosing a set of cities in this area and then determining the nodal structure that prevails. The nodal structure which emerges should resemble the known hinterland and ranking of the major cities in the area. Certain cities outside of the State were included in the study in order to examine the role they play in the network of city associations. Portland, Oregon and Vancouver, British Columbia were especially important additions.

The associations were defined by the number of long distance telephone messages between city-pairs during one week in June, 1958.[7] Certain cities were omitted from the study due to characteristics of the data and in order to limit the size of the study.

Many pairs of neighboring cities have direct dialing service and in these instances the intercity calls were not recorded in long distance data. Dormitory towns for Seattle and several "twin cities" such as Aberdeen-Hoquiam, Chehalis-Centralia and Pasco-Kennewick had direct-service exchange. This is not a serious deficiency in the data because such cities very likely function as a single point in the state-wide network, and one of the "twin cities" in each pair could be used in the study. Certain fairly large cities north of Seattle and along the Puget Sound were omitted for lack of data. These cities were serviced by a different telephone company. Because each year the telephone companies simul-

[6] For examples see, Luce, R.D. and D. Perry, "A Method of Matrix Analysis of Group Structure," *Psychometrika*, Vol. 14 (1949) pp. 95-116; and Katz, L., "A New Status Index Derived from Sociometric Analysis," *Psychometrika*, Vol. 18 (1953) pp. 39-44.

[7] We are indebted to the Pacific Bell Telephone and Telegraph Company and especially to Mr. Homer Moyer, a Seattle officer of that company, for this information.

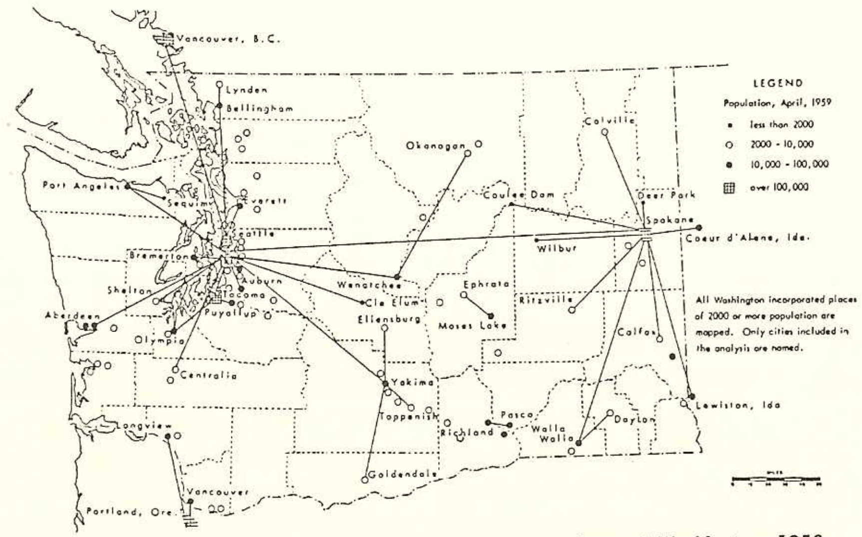


FIGURE 2. Nodal Structure Based on Telephone Data, State of Washington, 1958.

taneously take a one-week sample of intercity telephone calls, comparable data exist but there was no attempt to obtain them. Finally, all cities above a certain population size were not included in order to restrict the size of the study. Some small towns were chosen, however, in an effort to obtain samples of hierarchies with directed paths of length *2* or more. The map in Figure 2 identifies the cities in the study. With the advantage of hind-sight, it might have been preferable to have included more small towns in the study.

TABLE I. A Portion of the 40 × 40 Table of Number of Messages Between City-Pairs, for One Week of June 1958.

From City	To City								
	Code	01	02	03	04	08	13	26	27	40
Aberdeen	01	—	24	50	0	246	3671	54	4	1005
Auburn	02	26	—	35	0	8	7654	42	0	163
Bellingham	03	55	27	—	782	24	2494	101	3	356
Lynden	04	4	0	2250	—	4	357	9	0	110
Longview	08	329	15	32	0	—	1911	87	4	4773
⋮										⋮
Seattle	13	3427	4579	3843	308	1268	—	6168	269	16781
⋮										⋮
Spokane	26	61	32	119	6	85	9991	—	3842	3838
Couer d'Alene	27	0	4	4	0	6	254	5104	—	141
Portland, Ore.	40	802	210	304	22	4190	22179	3310	98	—

Largest column total—Seattle 154,192.

Certain channels of communication are omitted because they are not recorded in the long distance data. Direct-line calls are an example. Probably the only large volume, direct-line in the state links Seattle with Olympia, the state capital. If these data were included, the maximum association of Olympia might shift from Tacoma to Seattle. Though beyond the scope of this study, interesting results could be obtained if all channels of communication were included, such as radio, telegraph, mail and messenger service.

Table I is an example of the raw data tabulations. Forty cities were used in the study.[8] The entire table is the adjacency matrix of the almost completely connected graph of associations—there are a few zero entries. The row totals are the total out-contacts while the column totals are the total in-contacts. The direction of the message flow is read from the "row" city to the "column" city. The main diagonal entries are zero by convention.

TABLE II. A Portion of the Matrix B (Direct and Indirect Associations).

From City	To City									
	Code	01	02	03	04	08	13	26	27	40
Aberdeen	01	—	.248(4)	.395(4)	.551(6)	.166(3)	.245(2)*	.479(4)	.325(5)	.726(3)
Auburn	02	.303(4)	—	.364(4)	.109(5)	.108(4)	.508(2)*	.500(4)	.107(5)	.171(3)
Bellingham	03	.402(4)	.232(4)	—	.513(3)	.180(4)	.165(2)*	.739(4)	.247(5)	.254(3)
Lynden	04	.328(5)	.790(6)	.148(2)*	—	.307(5)	.239(3)	.716(5)	.689(7)	.753(4)
Longview	08	.221(3)	.150(4)	.252(4)	.341(6)	—	.131(2)	.699(4)	.325(5)	.316(2)*
⋮										⋮
Seattle**	13	.227(2)	.303(2)	.253(2)	.204(3)	.870(3)	—	.409(2)	.188(3)	.111(1)
⋮										⋮
Spokane	26	.568(4)	.421(4)	.953(4)	.536(5)	.688(4)	.649(2)*	—	.252(2)	.260(2)
Couer d'Alene	27	.650(6)	.332(5)	.340(5)	.560(7)	.459(5)	.191(3)	.335(2)*	—	.103(3)
Portland, Ore.**	40	.563(3)	.185(3)	.237(3)	.176(4)	.278(2)	.140(1)	.224(2)	.725(4)	—
Column Total		.548(2)	.588(2)	.613(2)	.866(3)	.585(2)	.102(0)	.229(1)	.311(2)	.563(1)

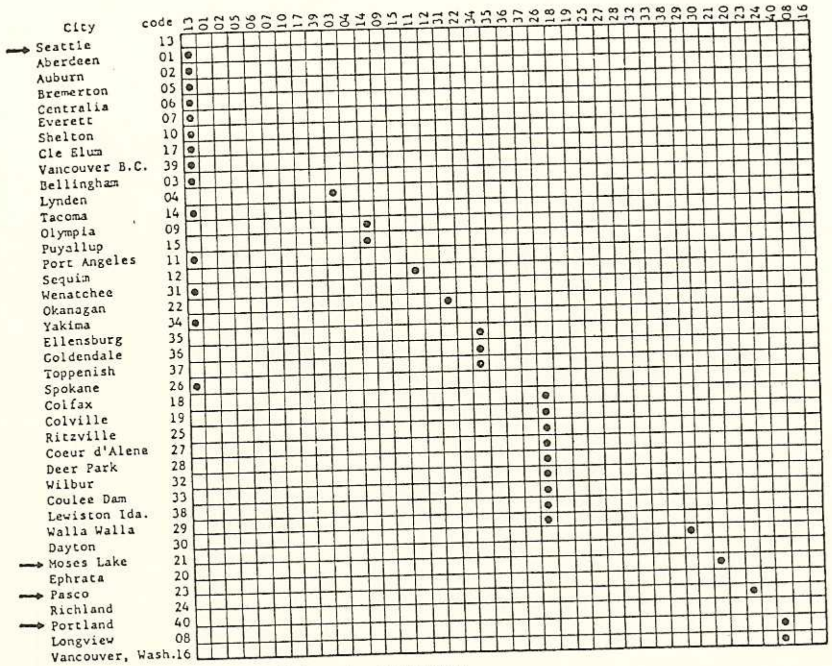
**Terminal point. *Nodal flow.

Remark: Figures are rounded to three significant digits. Data were processed to 8 significant figures. The value in parentheses represents the number of zeros before the first significant digit.

Table II is the adjacency matrix B which evaluates both the direct and indirect associations between the cities.[9] The nodal structure contained in this matrix was determined by (1) identifying the nodal flow, (2) ranking the cities by their total incoming associations (column totals), (3) assigning an orientation from cities with smaller total associations to one with a larger total association and (4) identifying the non-oriented cities as the center of its hierarchy. Figure 2 shows the results. Figure 3 is the adjacency matrix of the nodal structure derived from the direct and indirect associations.

[8] A copy of the entire matrix shown in Table I may be obtained from John D. Nystuen, Department of Geography, University of Michigan.

[9] The computations were made possible by a grant of computer time from Western Data Processing, University of California at Los Angeles.



Arrows indicate cities which are terminal points.

FIGURE 3. Matrix of the Nodal Structure.

CONCLUSIONS

The techniques defined in this paper will divide a set of cities into subgroups which specify a central place and its subordinate hierarchy. The association between cities is not the only system which may be defined as a network of points and lines. Nations or states may be thought of as points with migrations or commodity flows as lines. The important step in the employment of abstract linear graph analysis is the assignment of plausible meaning to the points and lines, preferably in terms of some real world phenomena. The usefulness of the attributes and the interpretation of the resulting hierarchy depends on the correspondence between an empirical example using graph theory analysis and other knowledge of the phenomena. The procedure described in this paper may be employed in a variety of ways, but the application is valid only when significant theoretical conclusions are produced and verified empirically.

Implications of the Nodal Structure in Washington State

The nodal regions that are suggested by the nodal structure agree, in general, with expectation. Seattle is the dominant center with nested hierarchies defined around Spokane and Yakima. Portland forms a system of its own by capturing nearby Washington State cities. The two small but independent hierarchies defined on Pasco and Moses Lake are most interesting.

Ephrata and Moses Lake are located on the boundary between two large hinterlands where it is postulated that self-reliance or independence is most likely to appear.[10] In addition, these two cities constitute an anomaly because, while Ephrata is an old city, Moses Lake was recently created by government fiat.

The small hierarchy with Pasco as the central point was anticipated by Ullman when he evaluated the growth centers of the western United States:[11]

> "One hypothesis that occurs to me for the future is that Pasco-Kenneick-Richland... might develop as the subregional shopping center, supplanting the dominance of older (and more attractive) Yakima and Walla Walla."

The effect of the national border is clear. Vancouver is subordinate to Seattle and it does not dominate any city in the study, even though it is a large city and is much nearer to Lynden and Bellingham than is Seattle. It is probable that Vancouver would have been a terminal point if other Canadian cities had been included within the study. This is not a defect in the method. The results are only an evaluation of the associations between the cities in the study.

The nodal region of Tacoma, a large city south of Seattle, is also anticipated by theory. Tacoma is dominant in a nearby region. This region is off-center, in the direction away from the larger city of Seattle. The dominance of Seattle re-asserts itself at even greater distances so that Aberdeen and Centralia are directly associated with Seattle, rather than by a two-link path through the closer and larger city of Tacoma[12] This and the other agreements with existing theory and accepted empirical evidence demonstrate the utility of the nodal structure for analyzing city associations.

Further Graph Theory Applications

Given a set of cities in an area and a measure of association between them, a set of hierarchies has been obtained. Even more information is desirable. Spokane is obviously the second most important central place in Washington state, yet it is subordinate to Seattle in a hierarchy while the much smaller places of Moses Lake and Pasco dominate their respective systems. Intuitively, a second in command position in a large organization is more important than the primary position in a tiny organization. Some measure of this difference in status is desirable. A further application of graph theory to this problem is suggested in a paper by Harary.[13] His ideas are adapted to this problem by the present authors in a further study of city associations.

---

[10] Hoover, E.M., *The Location of Economic Activity*, McGraw-Hill, New York, 1948. Isard, W., *Location and Space Economy*, John Wiley and Sons, New York, 1956. Lösch, A., *The Economics of Location*, Yale University, New Haven, 1954.

[11] Ullman, E.L., *Growth Centers of the West*, University of Washington, Seattle, 1955, p. 48.

[12] Hoover, E.M., *op. cit.* Isard, W., *op. cit.* Lösch, A., *op. cit.*

[13] Harary, F., "Status and Contrastatus," *Sociometry*, Vol. 22 (1959) pp. 23-43.

APPENDIX D

Previously unpublished telephone data (for the State of Washington) referred to in Appendix C, 1958—a starting point for a time-series. Nystuen and Dacey.

NUMBER OF TELEPHONE MESSAGES FOR ONE WEEK IN JUNE, 1958:
Selected towns in and near Washington State (40 x 40 matrix).

Towns by name and number \ Towns by number		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Aberdeen	01	---	24	50	0	90	196	93	246	740	204	116	9	3671	1519	44	84	3	0	0	5
Auburn	02	26	---	35	0	67	16	54	8	107	13	8	3	7654	2914	897	21	7	3	0	0
Bellingham	03	55	27	---	782	82	11	328	24	112	12	68	0	2494	446	22	28	0	0	0	4
Lynden	04	4	0	2250	---	8	4	39	4	4	0	4	4	357	27	6	0	4	0	0	0
Bremerton	05	102	59	90	4	---	40	188	40	347	287	271	26	13892	3089	67	44	3	4	0	4
Centralia	06	242	24	7	4	27	---	24	318	792	43	14	8	1167	799	36	55	4	0	0	0
Everett	07	74	50	671	38	160	21	---	62	206	35	92	9	8672	844	36	51	14	4	4	14
Longview	08	329	15	32	0	34	352	54	---	412	17	8	0	1911	720	17	988	0	4	0	4
Olympia	09	783	82	114	7	284	846	185	369	---	779	152	20	6162	6028	214	189	4	3	12	32
Shelton	10	263	20	6	0	255	64	32	19	875	---	65	6	1085	830	28	9	0	0	0	0
Port Angeles	11	108	6	66	5	209	16	100	10	131	54	---	1362	2945	466	9	19	0	0	0	0
Sequim	12	11	0	7	4	48	0	9	3	17	10	1535	---	393	55	0	5	0	0	0	0
Seattle	13	3427	4579	3843	308	8153	945	11455	1268	4989	956	2260	309	---	18445	2060	938	361	91	119	514
Tacoma	14	1247	2342	415	27	2856	604	938	607	5870	793	323	29	14450	---	11090	522	37	16	17	64
Puyallup	15	43	725	18	5	63	29	34	12	218	35	12	4	2534	14620	---	18	4	14	0	7
Vancouver, WA	16	63	8	18	4	39	37	39	908	182	10	12	4	1077	438	26	---	0	8	0	6
Cle Elum	17	7	7	0	0	5	0	20	0	0	0	4	4	546	49	6	2	---	0	3	13
Colfax	18	0	3	0	0	0	0	4	4	12	0	4	0	105	11	10	6	0	---	5	0
Colville	19	4	0	3	0	0	0	0	4	15	0	0	0	117	8	0	4	0	4	---	4
Ephrata	20	0	6	7	0	7	0	19	0	56	4	3	4	624	54	5	5	18	4	8	---
Moses Lake	21	6	12	13	0	11	3	33	5	24	0	3	0	1370	186	10	14	12	7	8	2294
Okanagan	22	0	9	4	0	0	4	5	4	12	0	0	0	219	15	0	0	0	0	4	43
Pasco	23	6	11	15	0	7	0	18	10	46	0	4	0	913	89	4	32	0	35	13	72
Richland	24	0	3	8	0	17	4	7	10	17	3	0	0	615	134	0	11	4	0	4	9
Ritzville	25	0	0	4	0	4	0	0	0	18	0	0	0	144	21	3	4	0	28	4	20
Spokane	26	61	30	119	6	101	18	196	85	364	27	45	7	9791	994	42	174	30	686	969	687
Coeur d'Alene ID	27	0	4	4	0	8	4	12	6	14	0	4	4	254	31	0	8	0	15	18	10
Deer Park	28	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	3	20	0
Walla Walla	29	6	10	16	4	12	3	41	34	62	9	4	0	1444	155	6	45	10	120	7	23
Dayton	30	0	0	0	0	0	0	4	6	12	0	0	0	58	10	9	13	0	0	4	0
Wenatchee	31	26	17	51	4	40	3	167	18	153	4	6	0	2713	9	9	24	52	7	8	468
Wilbur	32	0	0	0	0	0	0	0	4	3	0	0	0	35	5	0	0	0	0	9	44
Coulee Dam	33	6	4	4	0	0	3	5	0	5	0	0	0	109	13	0	0	0	3	9	72
Yakima	34	144	45	100	6	132	67	105	106	391	44	26	0	5763	895	45	111	186	19	19	237
Ellensberg	35	7	23	0	4	37	0	42	11	76	4	10	3	1282	189	20	19	0	4	6	64
Goldendale	36	0	4	12	0	7	0	0	0	11	0	0	0	65	10	0	0	0	0	4	0
Toppenish	37	0	6	9	0	5	0	6	8	11	0	0	4	369	46	0	6	3	0	0	5
Lewiston, ID	38	9	9	10	0	15	0	8	12	7	0	0	0	912	66	0	10	4	4	7	8
Vancouver, B.C.	39	31	0	604	47	27	11	1	23	43	18	42	0	5416	415	19	47	0	0	6	4
Portland, OR	40	802	210	304	22	374	353	721	4190	1004	128	182	24	21179	4439	140	36581	27	50	40	113

Data obtained by Michael F. Dacey, [then of] the University of Pennsylvania and John D. Nystuen, the University of Michigan. Courtesy of Homer Moyer, Pacific Bell Telephone Co., Seattle.

NUMBER OF TELEPHONE MESSAGES FOR ONE WEEK IN JUNE, 1958:
Selected towns in and near Washington State (40 x 40 matrix).

Towns by name and number \ Towns by number		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Aberdeen	01	9	7	4	0	0	54	4	0	21	0	23	0	0	67	17	4	4	15	22	1055
Auburn	02	22	0	13	10	0	42	0	0	6	0	10	0	0	53	23	4	3	4	4	163
Bellingham	03	15	0	18	5	0	101	3	0	18	4	55	3	4	60	13	6	7	7	591	356
Lynden	04	0	0	0	0	0	9	0	0	4	0	7	0	0	6	3	0	0	0	36	110
Bremerton	05	24	0	9	25	0	100	5	0	18	0	35	0	0	87	27	4	9	5	22	398
Centralia	06	0	0	0	0	0	13	4	0	6	0	0	0	0	31	4	0	0	0	7	510
Everett	07	26	6	13	11	0	135	11	0	31	0	145	3	0	92	27	3	10	8	3	798
Longview	08	12	0	12	11	4	87	4	0	25	4	13	0	0	88	8	0	12	12	28	4773
Olympia	09	34	9	17	17	16	267	3	0	67	4	89	3	0	229	44	5	14	7	50	1253
Shelton	10	0	0	0	0	0	16	0	0	6	3	5	0	0	14	5	0	0	3	9	165
Port Angeles	11	0	0	0	4	0	33	0	0	3	0	10	0	0	16	12	4	0	0	21	217
Sequim	12	4	0	0	0	0	6	0	0	3	0	0	0	0	7	0	0	0	0	0	52
Seattle	13	917	168	578	705	96	6168	269	19	1122	60	1886	35	94	4481	936	56	281	899	4851	16781
Tacoma	14	158	31	114	116	5	1096	27	3	182	6	6	7	11	758	170	9	48	37	392	5878
Puyallup	15	9	0	7	0	0	38	3	0	3	0	17	0	4	36	20	0	0	0	11	167
Vancouver, WA	16	14	0	13	13	4	125	0	4	37	4	16	0	0	94	15	0	10	9	50	37704
Cle Elum	17	8	4	0	4	0	30	0	0	23	0	74	0	0	238	0	0	7	0	4	57
Colfax	18	14	0	34	0	44	1162	12	6	145	17	3	7	4	14	4	0	4	0	0	123
Colville	19	7	4	3	0	4	1316	27	25	4	0	6	4	8	12	6	0	0	16	4	72
Ephrata	20	2004	40	55	17	41	1025	5	0	22	0	560	74	86	214	72	4	4	7	0	133
Moses Lake	21	---	27	166	41	107	1813	25	4	94	9	475	8	58	391	116	5	30	28	7	332
Okanagan	22	26	---	0	3	0	255	0	0	4	0	277	6	30	41	0	4	0	4	0	49
Pasco	23	156	4	---	5796	29	1236	12	4	1364	57	48	0	10	695	50	6	41	77	4	887
Richland	24	40	3	6799	---	4	489	18	0	341	13	25	4	0	379	19	4	24	16	5	345
Ritzville	25	106	154	52	12	---	119	16	3	34	5	16	7	4	47	10	10	0	10	3	77
Spokane	26	1164	157	800	461	627	---	3842	1295	1285	132	799	275	210	1147	134	19	62	1543	4	3838
Coeur d'Alene ID	27	18	3	17	5	6	5108	---	9	13	4	22	0	6	36	6	0	0	122	5	141
Deer Park	28	30	0	5	0	3	1563	12	---	4	0	0	0	0	4	0	0	0	3	0	21
Walla Walla	29	132	5	1082	373	26	1760	23	4	---	807	72	0	6	713	35	27	52	412	13	1467
Dayton	30	6	0	45	14	9	267	3	0	1171	---	5	0	0	51	9	0	3	57	0	83
Wenatchee	31	429	203	54	31	9	1167	28	4	66	5	---	16	75	735	0	3	11	26	30	499
Wilbur	32	9	4	3	0	8	446	4	0	4	0	23	---	189	8	0	0	4	3	0	6
Coulee Dam	33	47	34	11	0	6	385	9	0	0	4	43	345	---	16	0	0	0	3	4	29
Yakima	34	278	52	628	404	27	1366	33	0	711	44	663	22	38	---	1308	159	2494	50	3	1824
Ellensberg	35	137	9	43	18	8	191	0	3	35	6	0	3	4	1839	---	12	80	4	4	65
Goldendale	36	6	3	12	4	5	21	0	0	31	0	7	0	0	249	22	---	41	3	0	102
Toppenish	37	24	4	51	21	0	86	0	0	39	5	7	0	0	3698	97	36	---	4	4	231
Lewiston, ID	38	27	3	92	24	14	2577	112	10	438	56	27	0	5	103	7	0	13	---	7	561
Vancouver, B.C.	39	7	4	0	4	0	4	7	0	9	0	14	4	0	0	0	0	0	10	---	1384
Portland, OR	40	187	19	590	312	25	3310	98	21	503	50	315	16	22	1230	88	74	144	312	1362	---

Data obtained by Michael F. Dacey, Univ. of Pennsylvania and John D. Nystuen, Univ. of Michigan.
Courtesy of Homer Moyer, Pacific Bell Telephone Co., Seattle.

APPENDIX E

1. Comment by Nystuen, 1988;

2. Missouri map showing organizational hierarchy of telephone network, 1959—Nystuen and Dacey;

3. Previously unpublished telephone data (for the State of Missouri), 1959—another element or starting point for a time-series— Nystuen and Dacey.

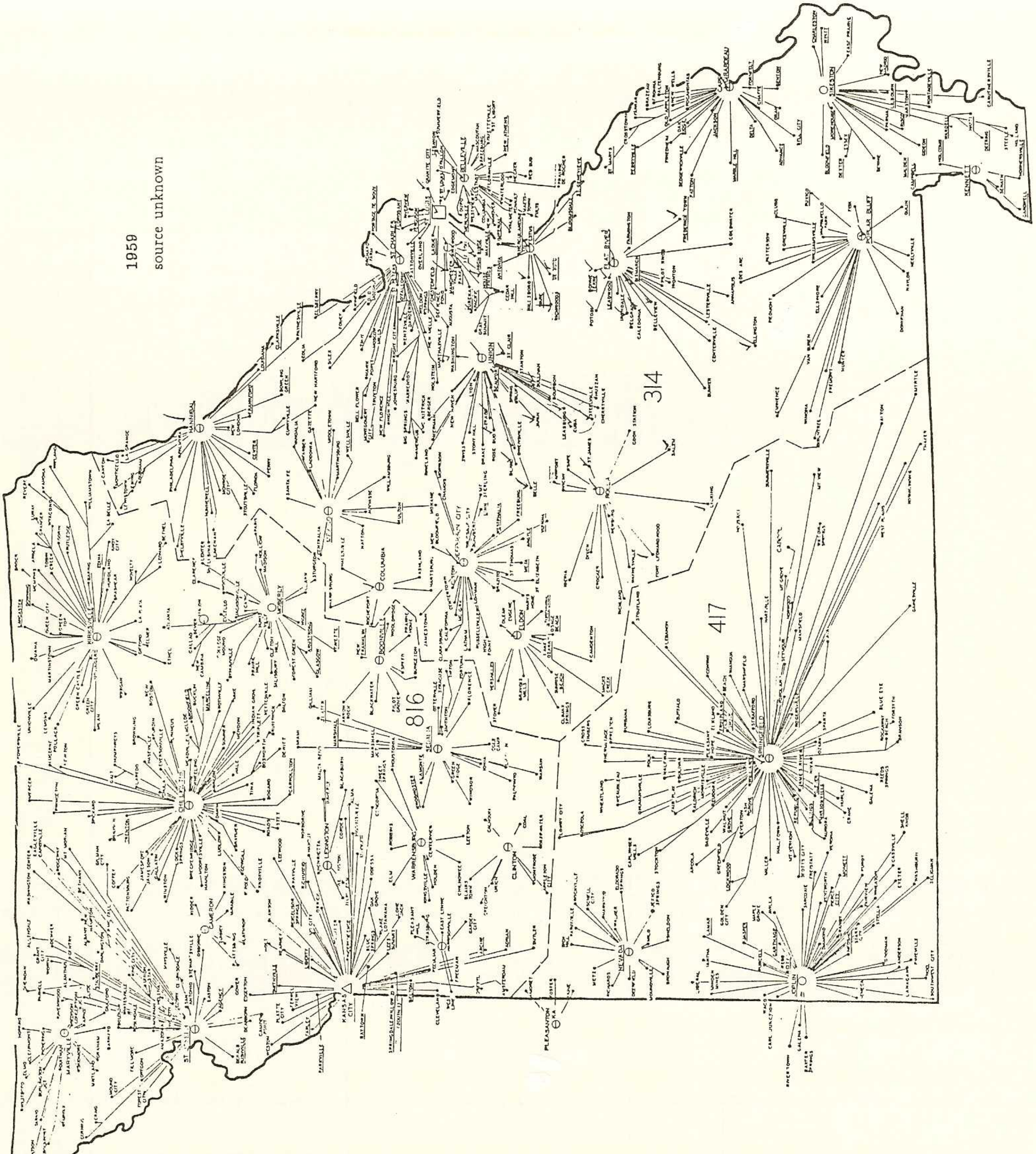
COMMENT BY NYSTUEN, 1988

The St. Louis message flow matrix displays the functional nodal region for that telephone system. The map displays the organizational hierarchy set up by the telephone company to handle this message network. An interesting question is "to what extent does the organizational system match the functional patterns in the message exchanges?" Tinkler's method can be applied to this question.

MISSOURI MAP SHOWING ORGANIZATIONAL HIERARCHY
OF TELEPHONE NETWORK, 1959

1959

source unknown



PREVIOUSLY UNPUBLISHED TELEPHONE DATA
(FOR THE STATE OF MISSOURI, 1959)—ANOTHER
ELEMENT OR STARTING POINT FOR A TIME-SERIES.

NUMBER OF TELEPHONE CALLS OCCURRING BETWEEN SELECTED PLACES IN MISSOURI OVER A TEN DAY PERIOD [1959].
Selected towns in and near Missouri (46 x 46 matrix).

TO: Towns by name and number	FROM: Towns by number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
St. Charles	01	---	9008	32	32	1238	78	14	0	4	2	4	61	20	16	10
St. Louis	02	7536	---	18866	17372	?	10640	3954	896	3750	898	1638	1042	744	725	615
Granite City, IL	03	54	23454	---	1336	410	38	16	6	14	10	2	3	2	9	11
Belleville	04	0	22012	1112	---	384	22	22	2	0	0	4	9	3	6	2
Kirkwood	05	410	?	370	400	---	5200	2198	316	816	110	206	116	91	99	36
Manchester	06	70	10644	32	10	4020	---	476	52	202	6	30	55	33	14	9
Valley Park	07	14	4234	14	6	1890	620	---	52	422	10	82	11	14	21	5
Pacific	08	6	1006	2	0	178	54	34	---	32	48	4	331	275	183	29
Fenton	09	16	3874	24	8	906	194	506	18	---	86	456	7	11	22	3
Cedar Hill	10	2	666	12	2	102	10	20	34	66	---	95	6	9	78	6
High Ridge	11	4	1056	4	6	250	34	90	2	494	144	---	2	3	20	4
Washington	12	44	1138	3	9	22	36	2	140	2	3	2	---	1167	251	70
Union	13	14	929	7	2	30	25	7	144	8	8	1	1156	---	709	100
St. Clare	14	22	839	6	7	83	13	10	115	11	58	33	354	840	---	357
Sullivan	15	10	809	11	14	16	9	8	22	4	2	8	95	138	394	---
Gerald	16	0	103	0	0	1	2	0	0	0	0	0	54	78	10	8
Rosebud	17	4	27	0	0	0	0	0	2	1	0	0	0	1	0	0
Bland	18	0	33	1	0	0	0	0	0	1	0	2	3	4	3	3
Owensville	19	1	243	1	3	9	0	0	5	0	0	1	88	45	19	33
Belle	20	1	87	1	1	1	2	0	1	0	0	1	10	10	6	0
Japan	21	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Cuba	22	2	443	5	7	10	7	2	4	3	4	0	12	17	44	98
Leasburg	23	1	40	0	0	1	1	0	0	0	0	0	1	1	1	40
Steelville	24	3	436	8	0	9	4	1	1	1	4	1	8	11	8	36
St. James	25	7	265	1	1	2	3	1	10	3	1	4	13	2	13	23
Rolla	26	10	1172	14	11	26	4	1	7	0	2	8	30	33	53	59
Ft. Leonard Wood	27	0	218	5	7	3	0	1	0	0	0	0	0	1	6	4
Licking	28	0	71	5	2	3	1	1	0	0	0	0	0	0	0	2
Salem	29	3	356	3	2	4	1	0	0	4	0	0	2	3	9	10
Flat River	30	10	1311	16	11	22	3	2	3	6	7	1	7	4	5	15
Bonne Terre	31	3	427	2	3	5	1	1	0	0	1	0	0	1	5	7
Potosi	32	3	968	21	4	18	3	3	2	2	1	0	1	8	11	24
Irondale	33	0	135	0	0	0	0	0	0	0	0	0	0	0	0	0
Caledonia	34	0	62	1	0	0	1	0	0	0	1	1	0	0	1	2
Bismark	35	0	224	2	2	3	0	0	0	0	0	0	0	0	1	0
Ellington	36	0	133	1	0	0	0	0	0	0	0	0	1	0	0	0
Frederick Town	37	8	795	12	0	0	0	2	1	0	0	1	0	1	6	1
Leadwood	38	1	147	0	0	1	1	0	0	0	1	0	0	0	1	7
Cape Girardeau	39	11	3234	27	44	6	3	1	4	3	2	4	4	8	3	0
Poplar Bluff	40	8	1571	28	13	6	0	1	0	4	0	3	1	2	5	0
Springfield	41	15	2634	23	28	68	6	6	3	5	2	1	8	12	10	20
Eldon	42	2	256	1	3	1	0	0	0	0	0	1	4	0	0	2
Lake Ozark	43	5	291	20	5	1	2	1	0	0	0	1	1	1	2	1

Jefferson City	44	36	2977	10	28	83	8	6	7	3	0	3	66	37	13	15
Columbia	45	54	2145	15	18	90	16	5	7	0	0	0	43	29	22	8
Hannibal	46	54	1451	8	14	20	2	0	2	2	0	1	8	4	7	1
COLUMN TOTALS		8445	101924	20726	19412	9922	17044	7392	1858	5863	1129	2599	3613	3663	2811	1676

Data obtained by Michael F. Dacey and John D. Nystuen.

NUMBER OF TELEPHONE CALLS OCCURRING BETWEEN SELECTED PLACES IN MISSOURI OVER A TEN DAY PERIOD [1959].
Selected towns in and near Missouri (46 x 46 matrix).

TO: \FROM: Towns by name and number	\Towns by number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
St. Charles	01	2	3	0	2	3	0	5	0	11	3	6	1	2	0	9
St. Louis	02	80	20	37	225	85	8	422	35	450	264	760	315	88	182	1106
Granite City, IL	03	7	0	1	1	1	0	7	2	8	6	19	16	2	13	17
Belleville	04	2	0	0	3	2	0	4	0	3	0	15	10	1	2	10
Kirkwood	05	3	0	2	11	1	1	18	0	26	7	78	4	9	12	50
Manchester	06	3	0	2	1	2	0	3	0	3	1	15	0	2	0	1
Valley Park	07	1	0	1	2	3	0	1	1	0	2	3	2	1	1	3
Pacific	08	2	0	0	5	0	0	1	0	8	14	14	1	0	1	3
Fenton	09	1	1	0	0	0	0	3	0	2	0	15	0	0	5	3
Cedar Hill	10	0	0	0	0	0	0	2	0	0	1	2	0	0	0	6
High Ridge	11	1	0	2	3	0	0	3	0	3	2	3	0	0	5	1
Washington	12	50	6	9	46	5	6	16	0	7	9	29	1	0	8	8
Union	13	78	5	4	41	11	1	10	0	17	2	28	1	1	2	9
St. Clare	14	7	0	1	16	1	6	46	6	15	12	39	2	0	10	10
Sullivan	15	30	0	9	37	2	0	158	55	50	19	86	2	6	0	12
Gerald	16	---	0	9	265	28	11	8	2	0	0	5	0	0	0	1
Rosebud	17	0	---	0	0	0	0	0	0	0	0	0	0	0	0	0
Bland	18	4	0	---	97	72	0	0	0	0	6	15	0	0	1	0
Owensville	19	137	44	153	---	100	5	46	0	12	13	25	0	0	4	1
Belle	20	15	7	165	155	---	0	3	0	3	0	0	0	0	9	1
Japan	21	0	0	0	0	0	---	0	0	0	0	1	0	0	0	0
Cuba	22	5	1	4	88	11	1	---	51	469	148	233	6	8	42	9
Leasburg	23	0	0	0	0	0	1	62	---	19	0	4	1	0	1	0
Steelville	24	1	0	0	18	1	1	572	21	---	65	101	1	1	117	23
St. James	25	2	0	12	20	25	1	133	1	56	---	858	16	10	26	3
Rolla	26	2	0	8	12	58	0	161	6	60	749	---	395	89	268	30
Ft. Leonard Wood	27	0	0	0	0	0	0	1	0	0	5	203	---	1	1	0
Licking	28	0	0	0	0	0	0	8	0	1	9	81	6	---	78	4
Salem	29	2	0	2	4	0	0	18	0	32	16	267	4	66	---	18
Flat River	30	0	1	0	1	0	0	7	0	12	4	25	4	1	16	---
Bonne Terre	31	0	0	0	0	0	0	1	0	29	1	2	0	0	4	1471
Potosi	32	0	0	0	2	1	0	4	0	26	0	15	0	1	6	292
Irondale	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	149
Caledonia	34	0	0	0	0	0	0	0	0	2	0	0	0	0	0	48
Bismark	35	0	0	0	0	0	0	5	0	0	0	6	0	0	8	633
Ellington	36	0	0	0	0	0	0	1	0	2	0	3	0	0	11	9
Frederick Town	37	0	0	0	0	0	0	1	0	5	0	3	0	0	7	259
Leadwood	38	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0
Cape Girardeau	39	1	0	0	1	1	0	2	1	1	1	20	2	4	11	127
Poplar Bluff	40	1	0	0	2	6	0	8	0	4	2	20	3	3	9	26
Springfield	41	0	0	1	6	2	0	9	0	10	27	222	180	27	46	14
Eldon	42	0	0	0	4	4	0	1	0	1	2	12	4	1	7	9

Lake Ozark	43	1	0	0	0	1	0	0	0	0	1	7	5	0	1	0
Jefferson City	44	6	3	16	32	70	0	17	0	26	32	283	169	6	39	13
Columbia	45	4	2	0	10	10	0	8	1	12	11	71	20	2	17	24
Hannibal	46	0	0	0	0	0	0	0	0	1	4	4	8	3	1	4
COLUMN TOTALS		448	93	438	1110	496	42	1775	182	1386	1439	3599	1179	335	973	4416

Data obtained by Michael F. Dacey and John D. Nystuen.

NUMBER OF TELEPHONE CALLS OCCURRING BETWEEN SELECTED PLACES IN MISSOURI OVER A TEN DAY PERIOD [1959].
Selected towns in and near Missouri (46 x 46 matrix).

TO: \FROM: \Towns by number / Towns by name and number		31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	ROW TOTALS
St. Charles	01	2	3	0	0	1	0	6	0	17	12	20	7	3	51	61	72	10821
St. Louis	02	549	603	60	37	234	76	539	0	1949	1148	1812	181	180	2739	1586	916	85362
Granite City, IL	03	10	25	3	0	7	1	13	0	29	18	26	5	5	9	21	10	25647
Belleville	04	3	1	0	0	0	1	0	0	38	31	17	4	3	19	20	17	23784
Kirkwood	05	26	17	0	0	12	2	18	0	179	21	289	13	11	139	99	52	11468
Manchester	06	4	5	0	0	0	0	3	0	1	3	11	2	11	19	18	3	15766
Valley Park	07	2	1	0	0	0	0	3	0	4	6	7	0	0	13	6	0	7458
Pacific	08	1	1	0	0	2	0	0	0	3	2	7	1	1	9	3	2	2263
Fenton	09	1	2	0	0	2	0	1	0	2	3	5	0	1	2	2	0	6182
Cedar Hill	10	0	2	0	0	0	0	0	0	1	1	2	2	1	0	1	0	1129
High Ridge	11	0	1	0	0	0	0	2	0	6	4	4	0	2	12	2	3	2172
Washington	12	0	4	0	0	1	0	2	0	8	0	8	3	1	60	34	8	3218
Union	13	1	9	0	0	0	0	3	0	12	6	8	2	7	56	25	10	3489
St. Clare	14	1	8	0	0	7	0	1	0	4	3	22	2	1	20	17	2	3007
Sullivan	15	4	25	0	1	0	0	1	0	1	0	39	0	1	17	10	2	2107
Gerald	16	0	0	0	0	0	0	0	0	0	1	1	0	2	11	1	0	601
Rosebud	17	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	45
Bland	18	0	0	0	0	0	0	0	0	0	0	1	1	0	22	2	1	272
Owensville	19	0	1	0	0	0	0	0	0	2	1	6	1	1	74	30	0	1104
Belle	20	0	0	0	0	0	0	0	0	0	3	7	6	1	0	7	1	504
Japan	21	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	5
Cuba	22	1	6	0	0	8	2	1	0	1	1	24	5	0	29	15	0	1827
Leasburg	23	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	175
Steelville	24	65	39	0	0	0	1	6	0	2	4	27	2	1	24	7	0	1617
St. James	25	1	0	0	0	1	0	2	0	0	2	44	1	1	36	16	1	1617
Rolla	26	2	13	0	0	4	0	2	0	19	24	353	14	7	282	98	8	4094
Ft. Leonard Wood	27	0	1	0	0	0	0	0	0	1	5	180	1	4	40	12	5	705
Licking	28	0	0	0	0	0	0	0	0	1	2	51	2	0	8	8	0	344
Salem	29	3	6	0	2	5	5	1	0	5	4	68	4	0	48	22	2	1001
Flat River	30	1747	244	120	28	533	4	232	0	182	40	15	3	0	17	26	5	4690
Bonne Terre	31	---	86	7	5	48	0	33	0	29	10	3	0	0	2	3	1	2191
Potosi	32	144	---	27	64	49	2	17	0	10	10	9	0	0	9	22	0	1779
Irondale	33	16	22	---	0	49	0	4	0	0	0	0	0	0	0	0	0	375
Caledonia	34	7	82	0	---	41	0	1	0	2	0	0	0	0	0	4	0	256
Bismark	35	42	29	39	33	---	2	10	0	17	7	1	0	0	1	1	0	1060
Ellington	36	1	3	0	0	6	---	4	0	10	33	17	0	0	5	11	2	253
Frederick Town	37	46	21	1	1	19	2	---	0	239	75	16	0	0	18	16	6	1562
Leadwood	38	152	76	0	0	0	1	20	---	5	1	0	0	0	0	0	2	459
Cape Girardeau	39	21	5	0	7	7	10	118	0	---	392	64	5	2	112	46	7	4322
Poplar Bluff	40	10	7	0	0	4	18	53	0	538	---	94	2	2	50	32	4	2541
Springfield	41	5	7	0	0	1	10	5	0	61	88	---	57	37	386	288	22	4352
Eldon	42	0	0	0	0	0	0	2	0	3	2	107	---	810	513	78	1	1831

Lake Ozark	43	0	0	0	0	2	0	0	0	2	2	82	1167	---	183	54	3	1842
Jefferson City	44	6	6	0	1	1	4	8	0	90	46	398	375	150	---	1628	103	6820
Columbia	45	7	12	0	0	3	6	15	0	39	23	280	61	27	1883	---	89	5089
Hannibal	46	0	0	0	0	0	1	1	0	4	1	19	2	2	125	83	---	1837
COLUMN TOTALS		2880	1373	297	179	1047	148	1127	0	3517	2035	4146	1931	1275	7053	4415	1360	

Data obtained by Michael F. Dacey and John D. Nystuen.

APPENDIX F

THE EXPECTED DISTRIBUTION OF NODALITY IN RANDOM (p,q) GRAPHS AND MULTIGRAPHS.

By

Keith J. Tinkler
Department of Geography
Brock University
St. Catharine's, Ontario

Paper presented to the East Lakes Association of American Geographers
October 15—16, 1976, Ann Arbor, Michigan.

ACKNOWLEDGMENTS

I am indebted to M. A. Hirst, D. C. Funnell and L. J. Wood for allowing me to use their network data, and to Alan Hay for his comments on an earlier draft of the paper.

NOTE

No originality is claimed for equations [1] and [2] as it is very probable that they have been derived before in the combinatorial or graph theory literature.

THE EXPECTED DISTRIBUTION OF NODALITY IN RANDOM (p,q) GRAPHS AND MULTIGRAPHS

One problem in studies of network structures represented as graphs is to find a suitable normative model to test an empirical structure against. One solution to this has been suggested in terms of structures defined as radial graphs (Tinkler, 1972). This paper suggests another approach using the frequency distribution of nodality in a graph. The test proposed is a graph counterpart to nearest neighbour models for points in Euclidean spaces. It enables a test to be made between an observed frequency distribution and the expected one for nodality in a graph with the same number of points and lines. The method is readily extendable to any number of nearest neighbours.

A (p,q) graph has p points and q lines. Its adjacency matrix, A, is $p \times p$ with main diagonal, $a_{ii} = 0$. When A has $(0,1)$ entries it defines a *graph* (Harary, 1969), and when it has $(0, integer)$ entries it defines a *multigraph*: i.e. links may be repeated between the same pair of points. For both types of graph the row sums of A define the degrees or nodalities, Δ, of the points of the graph G, $G = G(A)$. A frequency distribution, $f(\Delta, G)$ can be formed from these row sums. Half the sum of all the row sums defines the number of edges or lines, q, in the graph or multigraph. An expected frequency distribution of nodality is defined as $f^{\star}(\Delta, G)$ and arises from considering how the q lines are added to the p points. An urn model can be defined yielding the expected distributions for both graphs and multigraphs. The urn contains $P = \frac{1}{2}(p^2 - p)$ distinct *pairs* of labelled points. These represent all the possible lines in the graph. A random (p,q) graph is now obtained by making q draws from the urn, *without* replacing a pair that has been drawn. Further analysis of this model yields the hypergeometric distribution for which the probability that a point has degree exactly Δ :

$$prob(\Delta) = \frac{\binom{q}{\Delta}\binom{P-q}{p-1-\Delta}}{\binom{P}{p-1}}, \qquad [1]$$

$$P = \frac{1}{2}(p^2 - p), \qquad \Delta = 0, \ldots, p-1,$$

with mean $\frac{2q}{n}$ and variance

$$\left(\frac{2q(p-2)}{p^2}\right) \times \left(\frac{P-p}{p-1}\right).$$

The expected number of nodes of degree Δ is found by multiplying equation 1 by p, the number of nodes, and this yields for all the expected distributions, $f^{\star}(\Delta, G)$. The case for a multigraph, Gm, is simpler. The urn model permits the replacement of a drawn line and so the probabilities remain stationary with successive draws. The corresponding equation is:

$$prob(\Delta) = \frac{q!}{\Delta!\,(q-\Delta)!} \times \left(\frac{2}{p}\right)^{\Delta} \times \left(1 - \frac{2}{p}\right)^{q-\Delta}, \qquad [2]$$

$$\Delta = 0, \ldots, q,$$

with mean $\frac{2q}{n}$ and variance $\frac{2q(p-2)}{p^2}$. Again this is multiplied by p to get the expected number of nodes with this nodality. The calculation of [2] for all Δ yields $f^{\star}(\Delta, Gm)$. Obviously for any (p,q) graph the difference between $f(\Delta, G)$ and $f^{\star}(\Delta, G)$ can be used to see how the empirical graph departs from the expected structure. Two main departures from the expected can be envisaged. If there are a larger than expected number of points of high nodality then there is evidence of clustering of points in the network space so that a large number of points connect to a few points with high nodality. A corollary of this effect may be a larger than expected number of points with low nodality and since in the closed system an excess in one place implies a deficit elsewhere there may be a relative depletion in the centre of the expected distribution. This tendency is therefore one in which *clustering* is manifest and in which the differences are found towards the tails of the distribution. The extreme case is the star with one point of degree $p - 1$, and $p - 1$ points of degree 1. The opposite effect is when the number of nodes is larger than expected close to the average Δ, i.e. in the centre of the expected distribution. This type of departure implies *dispersion* of points in the network space and is a tendency towards a lattice in which each point has equal degree. A Christaller lattice, regular of degree 6, apart from points on the boundary in the finite case, would be an extreme example of this tendency. Similar remarks apply to the distributions for multigraphs and the interpretations are the same.

ASSUMPTIONS

A legitimate objection to these models has been pointed out to me by Alan Hay (personal communication). Both the above tests assume that the choices, the links being connected to a point, are made at random from the available set, whereas in practice this is unlikely. The implication is that *all* links in the system have an equal chance of being chosen, whereas distance decay notions suggest that short routes will be incorporated more frequently, and more probably, than long ones.

(i) The Binomial Case

The total number of draws and hence the mean is controlled by the p and q for the graph. Consequently the possible effect of the argument will be confined to the variance. In the binomial case the draws leave the probabilities unaffected and since the total weight of all the links to a point over those for all the rest of the system remains constant it is clear that the distribution of the weights on *the individual links* is irrelevant to a consideration of incidences at a point. The objection *is* relevant to the frequency distribution of the individual draws for particular links, but that is not the current problem. The binomial case is therefore immune to Hay's problem in this example. There is, however, another objection. In cases where the points of the graph are embedded in geographic space, (presume for simplicity the Euclidean plane), then the variable distances from a point to the other $(n - 1)$ points; a potential $(n - 1)$ links with weights, say, inversely proportional to their length, are not necessarily all from the same set. Points on the periphery will have a set of links with, on average, a greater length, and hence a total probability of being chosen lower, than for more centrally located points. In these circumstances the assumptions that the weight sets for every node are the same will not be true and there will be a bias in the results. Other things being equal the bias in the actual net would lead to a higher than expected number of low nodality places. This pattern is usually, but not always, associated with polarisation in the net so that this pattern may be *too* readily detected with reference to the binomial model which will tend to underestimate it.

(ii) The hypergeometric case

In this case each successive draw reduces the probability of another success for and particular node. When the links are weighted then the effect will be to reduce the variance in the distribution since the drawing of a *few* links becomes more probable, and of many, *less* probable: hence frequencies of very low degree and of very high degree become less probable and the variance about the mean is reduced. One deduces, therefore, that the test will be conservative for the polarised case, but biased for the lattice case: a lattice tendency may be too readily detected by comparison to the hypergeometric model.

The second objection: that weight sets for every node will not be identical again applies. The actual net in these circumstances would display more nodes of low degree, since their weight sets have fewer of the high probability, (short distance) links in them. Consequently the actual net, if it is built with distance links that are distance biased would *appear* more polarised than the probability model would suggest. Fortunately, the hypergeometric model applied to the weighted link case, is already conservative with respect to polarisation so that to some extent the two biases work against each other. We can expect, therefore, that the test will be reasonably robust.

STATISTICAL CONSIDERATIONS

Chi–squared and Kolmogorov–Smirnov tests

It seems natural to use a Kolmogorov–Smirnov one sample test to ascertain the goodness of fit between the observed and the expected distributions. This avoids the problems arising from the very large number of very small expected values encountered if a Chi–squared test is used. However, there are indications that even the K–S test is not ideal for this particular problem. The analysis is performed here for a six point star graph, Table I. Testing the maximum difference between the two distributions for N = 6, 0.399767 is not significant even at the 20% level. This is clearly unsatisfactory for a graph which has only a 1 in 3,000 chance of occurring under the urn model, and one which is extreme in terms of polarisation. The trouble partly stems from the fact that there is no suitable way of testing using a one-tailed test, the tabulated values are for two-tailed tests. However, the main problem arises from the fact that the observed frequency distributions are not drawn from an infinite population but are constrained by the total number of possible graphs for a given p and q. This is very limited for small graphs but grows rapidly, (see Appendix in Wilson, 1972). This appears to have the effect of making the test ultra-conservative. The effect might lessen for graphs with large p and q. It is also clear from the nature of the test that a few high nodality nodes do not themselves lead to large differences between the distributions and it is the effect these have on the frequency of low nodality nodes that mainly affects the size of the difference that is tested under H_0. There are then serious problems involved with either the Chi–squared test, or the Kolmogorov–Smirnov tests; in the latter case the effect of the finite population of observed distributions appears to have a conservative effect, but it is difficult to be certain. It is worthwhile, therefore, to propose an alternative method by which the tendency to clustering can be assessed.

EXACT PROBABILITIES

From the tabulated expected cumulative probability distribution one can identify points corresponding to the expected probability that the given (p,q) graph contains a point equal or larger than Δ. It is then only necessary to see whether the graph in fact contains a point of Δ equal to this or larger. This can be seen to be the case if the observed cumulative probability distribution is still smaller than the expected one. If it is less than the expected one then the H_0 that the graph does not contain such a point with a given significance level can be rejected with confidence equal to $1 - (cum.exp.Prob\,(\Delta))$. Applying this idea to Table I the expected probability of finding a node of degree 4 or larger is $(1.0 - 0.983016) = 0.16954$ and as the graph does contain such a node then H_0 is rejected with this significance. Similarly the probability that it contains a node of degree 5 is just 0.000333, and since the graph contains this node we can obviously reject H_0 very strongly.

This idea is useful for assessing the significance of observed nodes in the tails of the expected distribution, but is not applicable to deviations from the centre of the expected distribution. However, in these regions it is possible that the K–S test is helpful.

TENDENCY TO A NORMAL DISTRIBUTION

As n increases both the binomial and the hypergeometric cases tend to a normal distribution provided that the mean is far enough from the boundary of the distribution. In general this requires that for the values given the mean needs to be about 9 or more for the normal approximation to be useful. This necessity means that the approximation can only be applied to large and fairly well connected graphs: it is no use for cases 1,2,4,5,6,7 in Table 2. Using the normal approximation, when it is appropriate, a test of network structure is possible without extensive calculation. For example, using the weighted Kisii District Market visiting structure reported in Wood (1975) with p=80 and q=517 with $\Delta = 12.93$ the standard deviation is 3.55 so that with p=0.001 we should expect a node with nodality larger than 24. In fact a node with degree 77 exists in this multi–graph (Figure 1) so that we can say that the system is extremely polarised and very significantly so since it lies more than 18 standard deviations away from the mean!

EXAMPLES WITH ACTUAL NETWORKS

The observed distribution of nodality in a series of networks available to the author was tested against the expected distributions using equations [1] and [2]. Table 2 summarises the

TABLE 1:

Degree	Probability	Cum. Prob.	Obs. Prob. Cum.	Difference
0	0.083916	0.083916	0.000000	0.083916
1	0.349650	0.433566	0.833333	0.399767
2	0.399600	0.833166	0.833333	0.000167
3	0.149850	0.983016	0.833333	0.149683
4	0.016650	0.999666	0.833333	0.166333
5	0.000333	0.999999	1.000000	0.000001

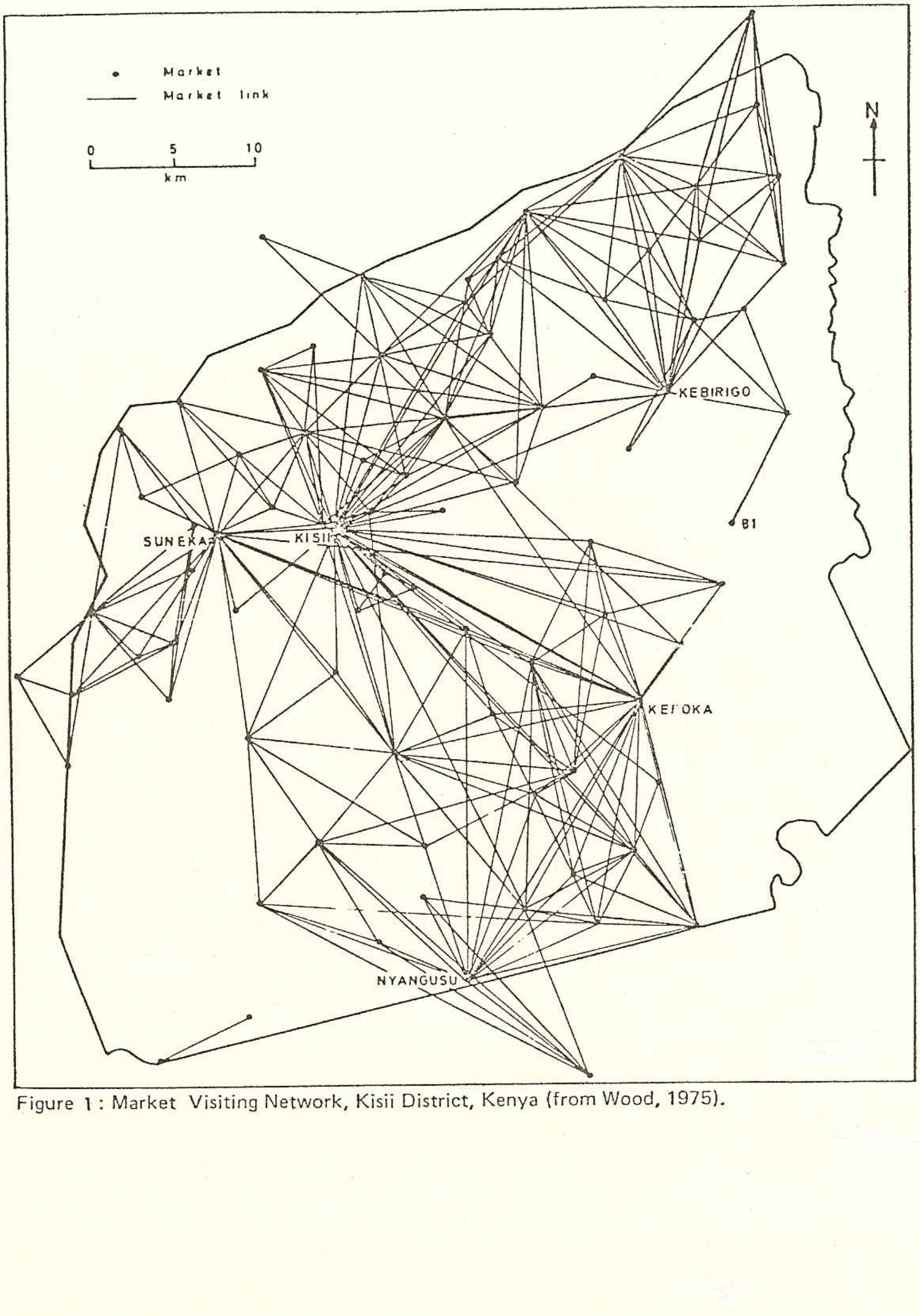


Figure 1 : Market Visiting Network, Kisii District, Kenya (from Wood, 1975).

TABLE 2:

ANALYSIS OF NETWORK STRUCTURES ON THE BASIS OF THE FREQUENCY DISTRIBUTION OF NODALITY.

	Network	Source	p	q	Δ	Sig.	Crit.	Diff.	Decision 1	Sig.	Decision 2
	GRAPHS										
1.	Uganda roads, 1921	Gould, 1967	18	17	1.8889	5%	.309	.1204	not sig.	0.5%	Δ > 5 sig
2.	Busoga roads, 1928	Hirst **	22	41	3.7273	5%	.294	.1283	not sig.	5.0%	Δ > 6 sig
3.	E. A. Airways, Domestic rtes. 1970	Hirst, Tinkler **	28	43	3.1071	5%	c.250	.2579	sig. cluster	0.0001%	Δ > 12 sig
4.	Uganda roads, 1935	Gould, 1967	31	44	2.8387	5%	.240	.0545	not sig.	2.0%	Δ > 6 sig
5.	Teso District, Uganda, Roads, 1970	Funnell 1972	45	64	2.8444	5%	.203	.0933	not sig.	2.0%	Δ > 6 sig
6.	Lanzarote, Canary Islands, roads.	Tinkler 1972	62	93	3.0000	5%	.173	.1196	not sig.	1.0%	Δ > 7 sig
7.	Sierra Leone, roads, 1970.	Harvey 1972	80	92	2.4500	5%	.152	.1158	not sig.	10.0%	Δ > 4 not sig
8.	Kisii District Markets, Kenya Market visiting structure	Wood 1975	80	301	7.5250	1%	.1825	.1977	sig. cluster	0.0001%	Δ > 21 sig
	MULTIGRAPHS										
9.	Teso District, Uganda, historical weights.	Funnell **	45	119	5.2889	1%	.1495	.2684	sig. cluster	0.03%	Δ > 14 sig
10.	Kisii District markets, usage weights.	Wood, 1975	80	517	12.9250	1%	.0715	.3835	sig. cluster	0.0001%	Δ > 32 sig

* In this column the significance level is the expected probability that there exists a node with nodality $\Delta > d$, where d is a nodal point determined by inspection from the observed distribution. When the decision is significant there are more observed nodes larger than than expected. The cumulated observed distribution is smaller at this point than is the cumulated expected distribution.

** Unpublished.

results; reasons of space prohibit a complete tabulation of the analysis for each network. As an example, a tabulation for an easily available network, Table 3, gives the analysis for the 1921 Uganda road network (Figure 2) published in Gould (1967) and Tinkler (1972). In Table 2 decision 1 is based on the Kolmogorov–Smirnov statistic, whereas decision 2 is based on the alternative exact probability method proposed above. The significance levels given are the best available and are rounded to the nearest convenient value since the expected distribution is a discrete one, not continuous. Decision 3 utilises the approximation to the normal distribution when it is appropriate. According to the K–S test only the airline network (Figure 3) and Wood's market visiting structure are significantly different from the expected distribution, and this was reasonable on the basis of the original graphs. For multigraphs, equation [2], both the graphs were significant. Funnell's multigraph representing the historical growth of routes in Teso District, Uganda, showed a marked deviance, as expected, although the $(0,1)$ structure of the same network was not significant by the same statistical test. On the other hand Wood's multigraph in which the $(0,1)$ structure of market visiting is weighed by the number of visits retains, and accentuates, the original deviation from the expected random structure.

In the tests based on the second criterion all but one of the networks turned out to differ from the expected distribution at the limits of the distribution. This can be interpreted to mean that every network contained at least one highly nodal element when the significance level was that given in the table. The only network that failed to have such an element was the Sierra Leone road network given by Harvey (1972) (Figure 4) and the network in fact is extremely dispersed. The graph theoretical dispersion was c.45,000.

Table 3 illustrates the complete working for the Uganda 1921 road network. The expected frequencies are seen to give a reasonable fit to the observed ones but the isolated node with degree 6 (Kampala in the actual network), has only a 5% chance of occurring in a random structure, and the cumulated observed probability distribution at $\Delta = 5$ is seen to be smaller than the expected one at this level. One may therefore conclude that the occurrence of this node is, with 99.5% confidence, not likely to be due to chance and indicates real polarisation tendencies in the network. Comparing the frequency distribution one may see the slight excess of terminal nodes, degree 1, and degree 2 nodes in the observed distribution, partly as a structural consequence of the hub created by the one node of degree 6.

n^{th} *ORDER NEAREST NEIGHBOUR ANALYSIS IN GRAPHS*

The adjacency matrix $A = A(G)$ represents the one step adjacencies in G. It is well known that if $a_{ii} = 0$ then A^n represents exactly n-step adjacencies in G. In general, however, A^n, has $(0, integer)$ entries and the graph of $A^n, G_n = G(A^n)$ yields, in general, a multigraph. Consequently two tests may be performed on G. It may be tested by equation (2) while it remains a multigraph, or one may apply a binary operator $b^\star$ to A^n such that $b^\star(a^n_{ij}) = 1$ if $a^n_{ij} > 0$, and 0 otherwise, and if $a^n_{ii} > 0$. Then the matrix $b^\star(A^n)$ can be tested by equation [1]. Logically the analysis can be carried as far as A^d where d is the topological diameter of the network. Beyond this point there are no points out of reach of any point in the network. However, for many points in the graph this will be larger than their König number and an alternative may be to carry the analysis only as far as r, where r is the radius of the network, the minimum König number.

An alternative procedure is to operate using $(A+I)^n$ since this will count all points up to and including a distance of n from each point; again this will generally be a multigraph even if the original graph is a graph and the operator $b^\star$ may be applied here too. It is worth noting that when $n = d$ then $b^\star(A+I)^n$ is simply a matrix full of ones, apart from the main diagonal which is set to zero in calculating the observed distribution to compare with the expected distribution computed from [1] or [2]. This means that $b^\star(A+I)^d = A(K_p)$ where K_p is the complete graph on p points and is trivially a lattice regular of degree $p-1$. These comments are also true if A, $(a_{ii} = 0)$, is a primitive matrix and d is equal to or larger than the solution number for the matrix.

These methods ought to be able to detect the presence of higher than first order clustering

TABLE 3:

UGANDA ROAD NETWORK 1921, AFTER GOULD 1967.

DEGREE	PROBABILITY	OBS. FREQ.	EXP. FREQ.	CUM. OBS. PROB.	CUM. EXP. PROB.	DIFFERENCE
0	0.119706	0.000000	2.154713	0.000000	0.119706	0.119706
1	0.288293	8.000000	5.189268	0.444444	0.407999	0.036445
2	0.304971	7.000000	5.489474	0.833333	0.712970	0.120364*
3	0.187482	2.000000	3.374677	0.944444	0.900452	0.043993
4	0.074688	0.000000	1.344383	0.944444	0.975140	0.030695
5	0.020358	0.000000	0.366453	0.944444	0.995498	0.051054
6	0.003909	1.000000	0.070359	1.000000	0.999407	0.000593
7	0.000536	0.000000	0.009652	1.000000	0.999943	0.000057
8	0.000053	0.000000	0.000950	1.000000	0.999996	0.000004
9	0.000004	0.000000	0.000067	1.000000	1.000000	0.000000
10	0.000000	0.000000	0.000003	1.000000	1.000000	0.000000
11	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
12	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
13	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
14	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
15	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
16	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000
17	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000

* Maximum difference

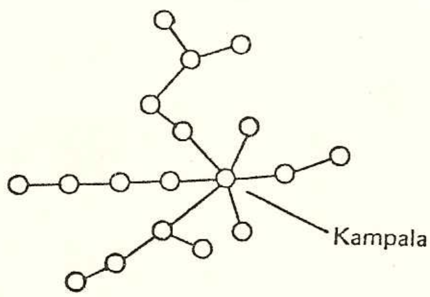


Figure 2: Uganda Roads 1921 (after Gould, 1967).

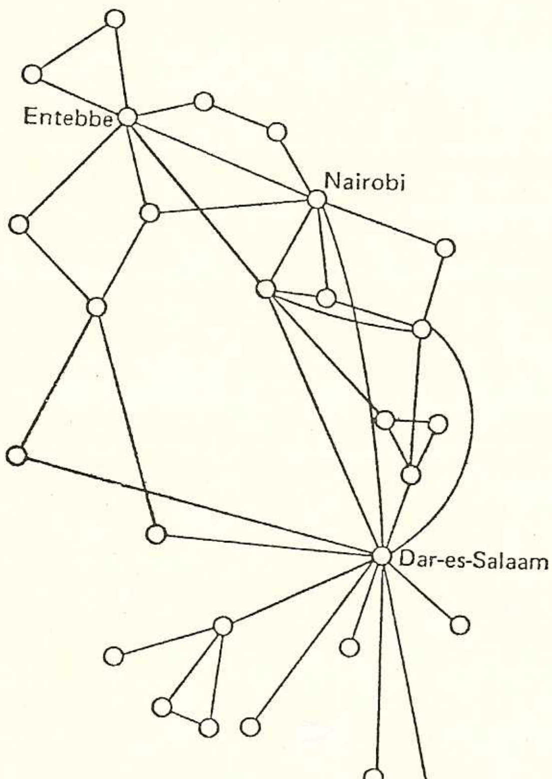


Figure 3: East African Airlines - Domestic Routes 1970.

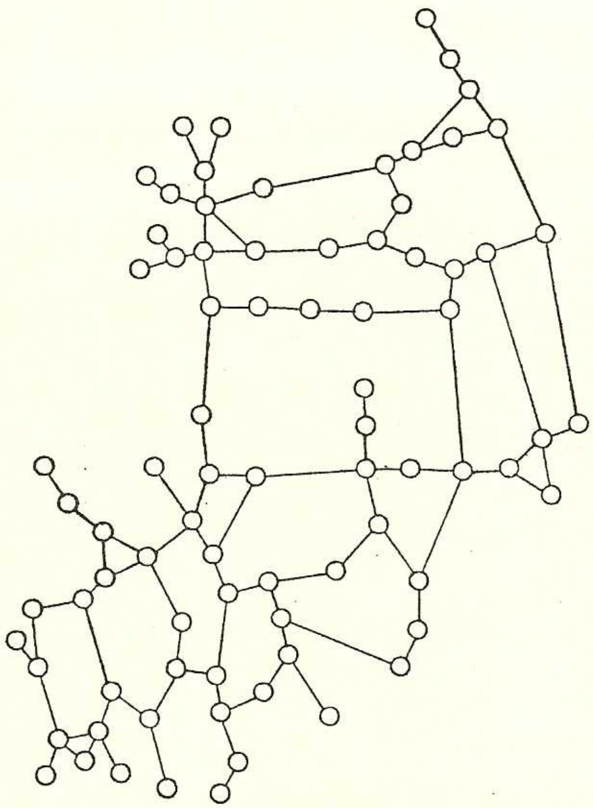


Figure 4: Sierra Leone Roads 1970 (simplified from Harvey, 1972).

in the graph. It has been shown in Tinkler (1972) that the vectors of row sums of successive matrix powers of A tend to the principal eigenvector of A and represent limiting accessibility in the network. Hence if equation [2] is applied to the powers of A without the operator $b^{\star}$ being applied then we have effectively devised a method of testing the nature of the frequency distribution of values on the principal eigenvector. Similarly the row sums of A itself are known to be exactly proportional to the fixed vector of the transition matrix of A, (Tinkler, 1973), and so the methods in this paper also test the significance of this vector in terms of its frequency distribution.

An n^{th} order example is given here for a simple network of 13 nodes and the computations were carried out up to $n = 4$ for the matrices $(A+I)^n$ with the main diagonal, $a - ii^n = 0$, and with the matrices subject to $b^{\star}$. A tabulation is given for $n = 3$ and a summary table for all n. The network is shown in Figure 5. The results are interesting because none of the networks display a significant difference on the K–S test which probably reflects the small N value and the limited nature of the possible observed distributions discussed above. This latter limitation becomes especially limiting for $b^{\star}(A^n)$ as n tends to d. This is because the number of possible (p,q) graphs reduces drastically as q approaches its limiting values of 0 and $\frac{1}{2}(p^2-p)$.

However, using the other criterion suggested it is possible to show that the graph does tend to deviate from the expected structure at its limits. This can be seen from Table 4 where it is clear that the structure possesses more *low* nodality nodes than expected. At the 4% level the observed value on the cumulated probability distribution is 0.153846 indicating a considerable surplus in this region. This is simply reversing the procedure used on Table 3. The high frequency of low nodality nodes carries the corollary of more than expected high nodality nodes, and this too can be seen from the Table. The corollary does not lead in this case to large differences between the cumulated probability distributions. It would seem that the criterion suggested is a useful one and emphasises which tendency is the more dominant element.

It is interesting that for n = 1, the original network, neither method indicates any differences from the expected distribution. This signifies that it may be necessary to carry out the analysis at higher n values to see if there are any significant advantages to be gained from connections within the graph at longer process lengths.

FREQUENCY DISTRIBUTION OF NODALITY IN DIGRAPHS AND MULTLIDIGRAPHS

The extension of the results in this paper to the more general case of digraphs, i.e. matrices of $(0,1)$ form but not symmetric and with main diagonal zero, and to multidigraphs, asymmetric $(0, integer)$ matrices is very easy. By substituting $\binom{1}{p}$ for $\binom{2}{p}$ in equation [2] and with q redefined as

$$\sum_{i \neq j} a_{ij},$$

we obtain the results for multidigraphs. By substituting $P = (p^2-p)$ for $P = \frac{1}{2}(p^2-p)$ in equation [1] we get the corresponding results for digraphs. In digraphs and multidigraphs the indegree of a point is usually different from the outdegree of the point so that in the adjacency matrix the row sums differ from the corresponding elements in the column sums. These results show that the frequency distribution of these elements should be the same irrespective of whether they are indegrees or outdegrees, although corresponding elements on the degree vectors may differ. By regarding any integral tally table as the adjacency matrix of a multidigraph one can use these methods to test the deviation of the frequency distributions of marginal totals from an expected norm.

CONCLUSIONS

A method has been suggested for the analysis of networks in terms of the frequency distribution of their nodality. It can be applied to graphs and multigraphs and is capable of indicating the extent to which there is a clustering or dispersion of nodes within the network

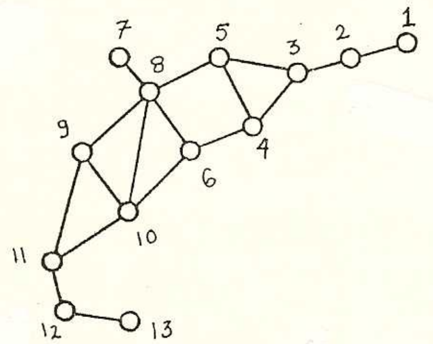


Figure 5: A simple test graph on 13 points.

TABLE 4:

THIRTEEN NODE TEST NETWORK, SEE FIGURE 1, ANALYSED FOR $b*(A + I)^3$ WITH ZERO MAIN DIAGONAL

Points: 13. Edges: 53. Mean nodality: 8.1538

DEGREE	PROBABILITY	OBS. FREQ.	EXP. FREQ.	CUM. OBS. PROB.	CUM. EXP. PROB.	DIFFERENCE
0	0.000000	0.000000	0.000002	0.000000	0.000000	0.000000
1	0.000005	0.000000	0.000071	0.000000	0.000006	0.000006
2	0.000104	0.000000	0.001348	0.000000	0.000109	0.000109
3	0.001102	0.000000	0.014325	0.000000	0.001211	0.001211
4	0.007292	2.000000	0.094800	0.153846	0.008504	0.145343
5	0.031762	0.000000	0.412907	0.153846	0.040266	0.113581
6	0.093615	2.000000	1.216989	0.307692	0.133880	0.173812
7	0.188566	0.000000	2.451363	0.307692	0.322447	0.014754
8	0.258156	1.000000	3.356033	0.384615	0.580603	0.195988
9	0.234688	2.000000	3.050939	0.538462	0.815291	0.276829*
10	0.134690	6.000000	1.750974	1.000000	0.949981	0.050019
11	0.043876	0.000000	0.570393	1.000000	0.993857	0.006143
12	0.006143	0.000000	0.079855	1.000000	1.000000	0.000000

* Maximum difference

TABLE 5:

NETWORK IN FIGURE 1 ANALYSED FOR FREQUENCY DISTRIBUTION OF NODALITY AT $n = 1$ TO 4 WITH MATRIX EQUAL TO $(b*(A + I)^{n})$ WITH ZERO MAIN DIAGONAL.

STEP	p	q		CRIT. VAL. 5%	MAX. DIFF.	DECISION 1		DECISION 2
1	13	17	2.6154	0.361	0.1015	not sig.		not sig.
2	13	37	5.6923	0.361	0.1797	not sig.	2%	2 excess low nodality
3	13	53	8.1538	0.361	0.2768	not sig.	4%	5 excess low nodality
4	13	64	9.8462	0.361	0.1480	not sig.	0.05%	6 excess low nodality

* Maximum difference

space. It is the graph counterpart of nearest neighbour analysis and can be carried out for n step adjacencies in the graph. Because the frequency distribution of nodality is linked with the frequency distribution of process in the graph via the theorem in Tinkler (1973) it provides a method by which this too can be tested in terms of the structure from which it arises. By taking the vector of row sums of any sufficiently high power of $(A+I)^n - I$ one has a test on the frequency distribution of accessibility in the original graph since it is known that this vector approaches the principal eigenvector of A as n tends to infinity, and the principal eigenvector indicates limiting accessibility in G (Tinkler, 1972).

It is hoped that these methods will provide readily implemented techniques for assessing empirical structures. The main drawback at this moment is a handy statistical test to compare the observed and expected frequency distributions. Refinements on, or alternatives to, the alternative criterion suggested here will be welcome. Geographically the notion of a randomly generated structure, and its characteristics, adds additional noise to normative theory! Classical theory would lead us to expect the widespread existence of lattice structures, dispersed networks. Conversely we can easily construct arguments that lead us to expect polarised structures especially in activity networks, such as Wood's market visiting network where the polarisation is evident even in the unweighted form, and in the unweighted structure of the East African Airways domestic air routes. In between these extremes we have the idea of a randomly constructed graph, arising from a multitude of local decisions often independent of the rest of a larger structure of which they eventually become part. The analyses given here suggest that several road networks tend to have these characteristics but that they may contain within them a number of nodes with a nodality that is improbable under the postulated urn models.

REFERENCES

Gould, P. R., 1967, "On the Geographical Interpretation of Eigenvalues," *Transactions, Institute of British Geographers*, No. 42, pp. 53-86.

Funnell, D.C., 1972, "Service Centres in Teso," *Occasional Paper No. 46*, Department of Geography, Makerere University, Kampala.

Harary, F., 1969, *Graph Theory*, Addison-Wesley.

Harvey, M. E., 1972, "The Identification of Development Regions in Developing Countries," *Economic Geography*, 48, No. 3, pp. 229-243.

Tinkler, K. J., 1972 (a), "Bounded Planar Networks: a Theory of Radial Structures," *Geographical Analysis*, 4, No. 1, pp. 5-33.

Tinkler, K. J., 1972 (b), "The Physical Interpretation of Eigenfunctions of Dichotomous Matrices," *Transactions, Institute of British Geographers*, No. 55, pp. 17-46.

Tinkler, K. J., 1973, "Letter to the Editor," *Transactions, Institute of British Geographers*, No. 58, pp. 149-150.

Wilson, R. J., 1972, *Introduction to Graph Theory*, Oliver and Boyd, Edinburgh.

Wood, L. J., 1975, "The Functional Structure of a Rural Market System," *Geografiska Annaler*, 57B, 2, pp. 109-118.

APPENDIX G

"Advice to authors" and other fliers. Institute of Mathematical Geography.

INSTITUTE OF MATHEMATICAL GEOGRAPHY
2790 Briarcliff
Ann Arbor, MI 48105

IMaGe MONOGRAPH SERIES–ADVICE TO AUTHORS

ADVICE TO AUTHORS

Manuscripts should be submitted directly to the editor. The monographs in this series are produced on demand from camera-ready copy supplied by the author. All manuscripts will be reviewed by at least two members of the editorial board or by referees suggested by them. Initial submissions should be typed, double-spaced, with one inch margins on all sides. Submit three copies, at least one of which should be copied on one side only. Figures should be in close-to-final form. Once a manuscript has been accepted, the author will be provided with further suggestions for the preparation of the final document. Authors should retain copies of all submissions as the high cost of postage prohibits the return of manuscripts. Royalty checks will be issued periodically, depending on sales. There is no charge to the author for producing these books. Before a book is released, authors will be provided with a transfer of copyright form, and copyright will be procured for each volume in the series by IMaGe.

PURCHASE OF MONOGRAPHS

Monographs may be ordered directly from IMaGe; the price of a monograph does not include postage and handling. A list of monograph titles and abstracts appears on the last page of each book. Because these books are published on demand, they do not go out of print. Standing orders, as well as individual orders, are welcome; the former well be filled ahead of the latter.

EDITORIAL BOARD

INSTITUTE OF MATHEMATICAL GEOGRAPHY
2790 Briarcliff
Ann Arbor, MI 48105

IMaGe MONOGRAPH SERIES–1988 PRICE LIST (EXCLUSIVE OF SHIPPING)

1. Sandra L. Arlinghaus and John D. Nystuen. Mathematical Geography and Global Art: the Mathematics of David Barr's "Four Corners Project," 1986. $9.95.

This monograph contains Nystuen's calculations, actually used by Barr to position his abstract tetrahedral sculpture within the earth. Placement of the sculpture vertices in Easter Island, South Africa, Greenland, and Indonesia was chronicled in film by The Archives of American Art for The Smithsonian Institution. In addition to the archival material, this monograph also contains Arlinghaus's solutions to broader theoretical questions–was Barr's choice of a tetrahedron unique within his initial constraints, and, within the set of Platonic solids?

2. Sandra L. Arlinghaus. Down the Mail Tubes: the Pressured Postal Era, 1853-1984, 1986. $9.95.

The history of the pneumatic post, in Europe and in the United States, is examined for the lessons it might offer to the technological scenes of the late twentieth century. As Sylvia L. Thrupp, Alice Freeman Palmer Professor Emeritus of History, The University of Michigan, commented in her review of this work "Such brief comment does far less than justice to the intelligence and the stimulating quality of the author's writing, or to the breadth of her reading. The detail of her accounts of the interest of American private enterprise, in New York and other large cities on this continent, in pushing for construction of large tubes in systems to be leased to the government, brings out contrast between American and European views of how the new technology should be managed. This and many other sections of the monograph will set readers on new tracks of thought."

3. Sandra L. Arlinghaus. Essays on Mathematical Geography, 1986 $15.95

A collection of essays intended to show the range of power in applying pure mathematics to human systems. There are two types of essay: those which employ traditional mathematical proof, and those which do not. As mathematical proof may itself be regarded as art, the former style of essay might represent "traditional" art, and the latter, "surrealist" art. Essay titles are: "The well-tempered map projection," "Antipodal graphs," "Analogue clocks," "Steiner transformations," "Concavity and urban settlement patterns," "Measuring the vertical city," "Fad and permanence in human systems," "Topological exploration in geography," "A space for thought," and "Chaos in human systems–the Heine-Borel Theorem."

4. Robert F. Austin, A Historical Gazetteer of Southeast Asia, 1986. $12.95.

Dr. Austin's Gazetteer draws geographic coordinates of Southeast Asian place-names together with references to these place-names as they have appeared in historical and literary documents. This book is of obvious use to historians and to historical geographers specializing in Southeast Asia. At a deeper level, it might serve as a valuable source in establishing place-name linkages which have remained previously unnoticed, in documents describing trade or other communications connections, because of variation in place-name nomenclature.

5. Sandra L. Arlinghaus, Essays on Mathematical Geography–II, 1987. $12.95

Written in the same format as IMaGe Monograph #3, this volume contains the following material: "Frontispiece–the Atlantic Drainage Tree," "Getting a Handel on Water-Graphs," "Terror in Transit: A Graph Theoretic Approach to the Passive Defense of Urban Networks," "Terrae Antipodum," "Urban Inversion," "Fractals: Constructions, Speculations, and Concepts," "Solar Woks," "A Pneumatic Postal Plan: The Chambered Interchange and ZIPPR Code," "Endpiece."

6. Pierre Hanjoul, Hubert Beguin, and Jean-Claude Thill, Theoretical Market Areas Under Euclidean Distance, 1988. (English language text; Abstracts written in French and in English.) $15.95.

Though already initiated by Rau in 1841, the economic theory of the shape of two-dimensional market areas has long remained concerned with a representation of transportation costs as linear in distance. In the general gravity model, to which the theory also applies, this corresponds to a decreasing exponential function of distance deterrence. Other transportation cost and distance deterrence functions also appear in the literature, however. They have not always been considered from the viewpoint of the shape of the market areas they generate, and their disparity asks the question whether other types of functions would not be worth being investigated. There is thus a need for a general theory of market areas: the present work aims at filling this gap, in the case of a duopoly competing inside the Euclidean plane endowed with Euclidean distance.

(Bien qu'ébauchée par Rau dès 1841, la théorie économique de la forme des aires de marché planaires s'est longtemps contentée de l'hypothèse de coûts de transport proportionnels à la distance. Dans le modèle gravitaire généralisé, auquel on peut étendre cette théorie, ceci correspond au choix d'une exponentielle décroissante comme fonction de dissuasion de la distance. D'autres fonctions de coût de transport ou de dissuasion de la distance apparaissent cependant dans la littérature. La forme des aires de marché qu'elles engendrent n'a pas toujours été étudiée ; par ailleurs, leur variété amène à se demander si d'autres fonctions encore ne mériteraient pas d'être examinées. Il paraît donc utile de disposer d'une théorie générale des aires de marché : ce à quoi s'attache ce travail en cas de duopole, dans le cadre du plan euclidien muni d'une distance euclidienne.)

INSTITUTE OF MATHEMATICAL GEOGRAPHY
2790 Briarcliff
Ann Arbor, MI 48105

IMaGe MONOGRAPH SERIES--INFORMATION VALID TO 1/1/89

UNIVERSITY LIBRARY STATEMENT

IMaGe Monographs are produced on demand from camera-ready copy, supplied by the author. Guidelines for authors, and a statement of IMaGe review policy appear on a different flier. Price structure is negotiated with Michigan Document Services, Inc., on a regular basis, reflecting changing costs in producing books. Fluctuations might be up or down--up, for example, if equipment maintenance costs rise; down, for example, if the factory location is moved to a cheaper site. Because IMaGe is responsive to the concerns of Acquisitions Departments in university libraries, it has established the following guidelines concerning the frequency of publication of monographs.

1. No single book shall be of a length forcing its cost to exceed $20.00, exclusive of postage and handling.

2. No more than six books shall be issued in a calendar year. If a single monograph number is of a length requiring more than one book for publication, then each of those books counts as one of the six. Therefore, the maximum amount of a basic annual subscription is $120.00.

3. Libraries should specify, at the time the order is placed, the preferred mode of shipping; if left unspecified, books will be shipped via Fourth Class mail, or Library Rate (whatever applies), in the continental United States, and via Air Mail (using a mail company to ascertain the the cheapest mode of transmission) to overseas and foreign libraries.

INSTITUTE OF MATHEMATICAL GEOGRAPHY
2790 Briarcliff
Ann Arbor, MI 48105

MICMG DISCUSSION PAPERS, JOHN D. NYSTUEN, EDITOR

The collected work of the Michigan Interuniversity Community of Mathematical Geographers is reprinted in a single essay format (as they were originally reproduced). Royalties from sales will be deposited in the "MICMG" fund for the development of projects in mathematical geography, administered by IMaGe, and disposed of by Nystuen, in collaboration with IMaGe.

Consider ordering one number as reading supplementary to texts in an upper division course. The dates or original release and titles of the individual numbers are listed below.

1. Arthur Getis, Temporal land use pattern analysis the use of nearest neighbor and quadrat methods. July, 1963.

2. Marc Anderson, A working bibliography of mathematical geography. September, 1963.

3. William Bunge, Patterns of location. February, 1964.

4. Michael F. Dacey, Imperfections in the uniform plane. June, 1964.

5. Robert S. Yuill, A simulation study of barrier effects in spatial diffusion problems. April, 1965.

6. William Warntz, A note on surfaces and paths and applications to geographical problems. May, 1965.

7. Stig Nordbeck, The law of allometric growth. June, 1965.

8. Waldo R. Tobler, Numerical map generalization; and Notes on the analysis of geographical distributions.

9. Peter R. Gould, On mental maps. September, 1966.

10. John D. Nystuen, Effects of boundary shape and the concept of local convexity; Julian Perkal, On the length of empirical curves; and, Julian Perkal, An attempt at objective generalization. December, 1966.

11. E. Casetti and R. K. Semple, A method for the stepwise separation of spatial trends. April, 1968.

12. W. Bunge, R. Guyot, A. Karlin, R. Martin, W. Pattison, W. Tobler, S. Toulmin, and W. Warntz, The philosophy of maps. June, 1968.